LABX CENTRO PARA A INOVAÇÃO NO SETOR PÚBLICO

ama AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA

Citizen-centric artificial intelligence

A systemic approach to virtual assistant evaluation in public services

February 2021

REPÚBLICA PORTUGUESA
Cofinanciado por:
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Social Europeu

Index

Presentation structure

Framework

LabX was challenged to evaluate the performance of the Virtual Assistant (VA) of the Tax and Customs Authority (AT) portal, which was running in a pilot phase, in order to identify a set of improvement recommendations to optimize the VA (*chatbot*).

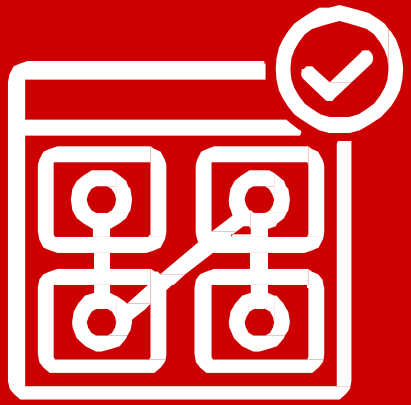
In order to carry out the assessment of the VA, which uses artificial intelligence, an approach was designed that is able to fully capture the different facets associated with this type of technology, which combines the analysis of the technology, in terms of its level of readiness, with the analysis of the assessment carried out by users and the study of the interaction between humans and machines.

This unprecedented approach makes possible both an internal evaluation, made by the VA host entity itself, how it sees itself and the relationship of the VA with the citizens, and an external one, made by the citizens, how they see the VA and what is their feedback on the relationship with the VA. In this way, through a systemic approach (with several exploitation routes), and by giving voice to citizens, we tried to contribute to focus this type of technologies on their needs and expectations.

This project considered two main purposes:

- inform recommendations for improvement to the VA concerned through the results of the application of this approach;
- validate, with a real case, the approach itself and the instruments developed for assessing VA performance in public services.

To accomplish this project, LabX established two partnerships with two specialized centres, one of them in emerging technologies (DTx), and the other in behavioural sciences (Behavioral Research Unit).



2.METHODOLOGICAL PROPOSAL

Methodological proposal

Exploitation routes and partnerships

Multidimensional readiness

The readiness assessment of the technological solution allows characterising its current technical properties according to solid guidelines and bringing clarity to the development programme that is open to this type of solutions. On the other hand, extending the readiness assessment (in an internal view) to other domains (organisational, social and legal), inseparable from the technological facet, makes it possible to have a view in several dimensions and thus potentiate its development.

Users' needs, expectations and “pains”

The definition of priorities and directions for the development of a technological solution should be based on a robust and up-to-date knowledge of the needs, expectations and "pains" felt by users.

Interaction between humans and machines

To complement the rigorous measurement of the solution's technical performance, it was important to ensure an evaluation of the interaction with the VA. Obtaining feedback on user satisfaction allows situating this solution from its tangible consequences and characterising its impact on its users.

Methodological proposal

Exploitation routes and partnerships

Virtual assistant metrics

The information produced by the *chatbot* support platform provides important data to complement some of the aspects of its interaction with citizens (no. of referrals to other channels when it is unable to respond to what is requested, for example) and, fundamentally, the intensity of use (no. of active users, no. of messages exchanged, for example). For this reason, it is very important to be able to count on these metrics to strengthen the assessment of the performance of the VA.

Partnerships with the scientific system

To carry out this project, it was essential to expand the execution capacity and bring on board the necessary technical expertise required. To this end, two partnerships were developed with two major research centres in their respective fields, ensuring open innovation and the activation of transfers with the innovation ecosystem in order to bring improvements to public services.

Methodological proposal

A systemic view of exploitation routes

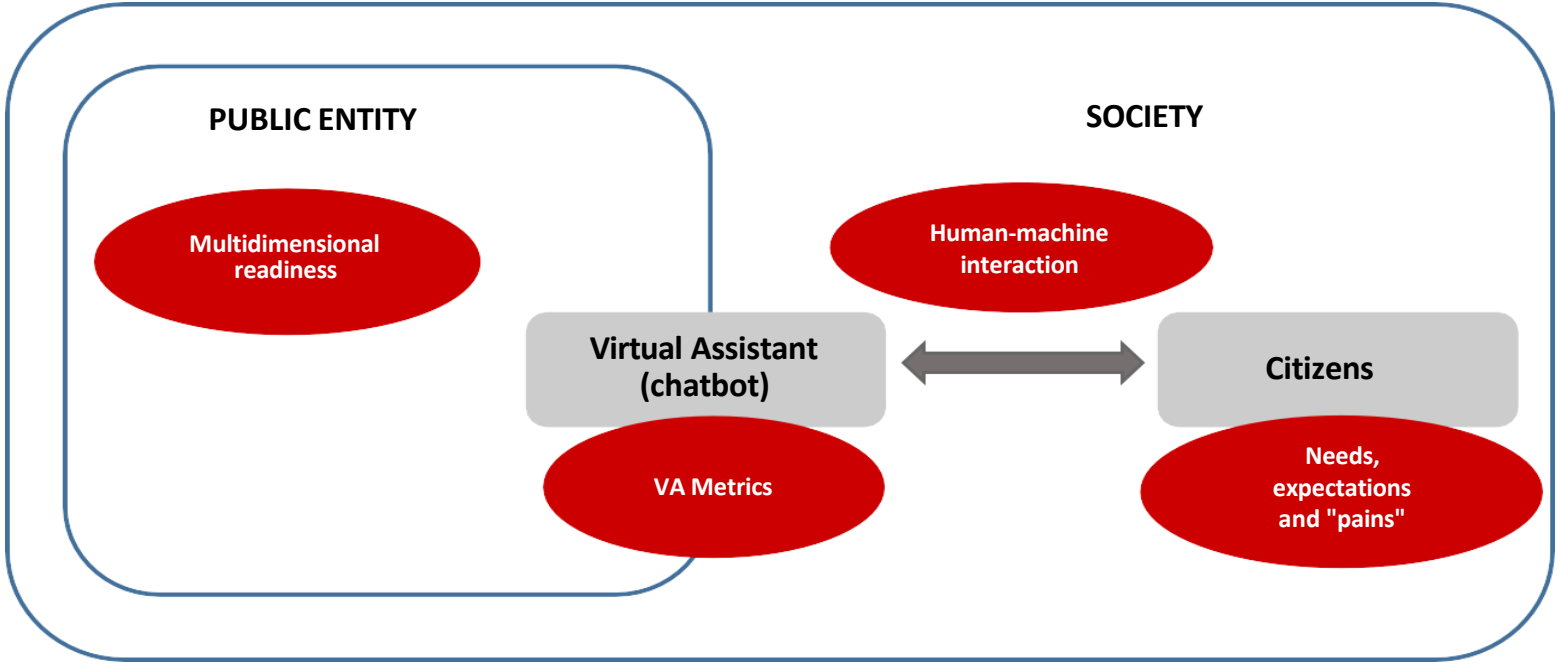


Image 1. Methodological proposal for a systemic vision

3.MULTIDIMENSIONAL READINESS

Multidimensional readiness

VA Solutions Maturity Assessment Tool

Based on a proposed evaluation model that has been developed by the European Commission[1] with the purpose of assessing the readiness levels (Technology Readiness Level) of an artificial intelligence solution applied to public services, a version was developed - in collaboration with partner DTx – to be adapted to the Portuguese context in order to allow the assessment of the level or stage of maturity of technological solutions through a multidimensional way.

In this sense, since artificial intelligence does neither appear from nowhere, having an associated purpose, nor develops in absolute isolation, we applied an original model that articulates four domains, characteristic of the public sector: **Technological (artificial intelligence)**, **Social**, **Legal** and **Organizational**.

Thanks to these results, we not only have a vision of the various dimensions of the technology, without suffering from "technocentrism", but we can also define strategies for this technology that take advantage of the latent opportunities in all these components, being aware of the overlapping obstacles that may arise between them.

This exploitation route allows for a self-diagnosis on the status of the adopted VA solution and enables the entity to have a strategic vision on the VA (in the legal, organisational, technological and social domains) and, based on the self-diagnosis, allow the definition of a strategy to be adopted for its optimisation and incremental improvement.

(1) Bruno et al.(2019) Technology Readiness revisited: A proposal for extending the scope of impact assessment of European public services. Accessible at https://ec.europa.eu/isa2/sites/isa/files/technology_readiness_revisited_-_icegov2020.pdf

Multidimensional readiness

The indicators

For each domain, three indicators were considered to characterise the different properties and features of the adopted solution (language processing tool, existence of teams with the skills to manage the VA, value assigned by citizens to the solution, regulatory framework, for example), with different impacts on maturity and its stage of development. The indicators served as a basis for the construction of a self-diagnosis questionnaire.

Artificial Intelligence domain

The choice of indicators was based on the fundamental characteristics that, for this domain, should be taken into account to maintain a fluid conversation and quality dialogue with users.

- Conversation tools - Aims to know which tools are adopted to enrich the content of the conversation and ensure greater robustness in actions and maintaining dialogue;
- Language Processing - Identifies a set of methodologies and tools directly applicable to the content of the intelligent system input so that the detection of intent and the selection of responses are the most appropriate for the interaction;
- Conversation methodology - It characterises the core of the intelligent system. The conversation methodology is responsible for the user-VA interaction (detecting intentions and generating a response to the user).

Multidimensional readiness

The indicators

Social domain

The VA should meet the needs of the target audience, so the self-diagnosis should assess the impact and value it delivers to users (the utility), what experience it provides (its quality - usability), and what barriers it poses. For this domain, the following indicators were considered:

- Basic requirements - Considers and takes into account the essential properties of the VA in meeting the needs of citizens and in promoting their adoption (ease of use and accessibility for all citizens);
- Social performance - For this indicator, issues related to cognitive effort and usability aspects (that the experience provides) are observed;
- Social barriers - This indicator takes into account aspects of fairness, transparency and trust that the VA transmits to citizens;

Multidimensional readiness

The indicators

Legal Domain

In the development and implementation of a VA it is necessary to ensure its compliance with the regulatory system and simultaneously consider ethical and privacy aspects. The following indicators have been associated with this domain:

- Data privacy - This indicator focuses on the privacy of citizens' data generated and exchanged during interaction with the VA. It identifies approaches to ensure the privacy of such data;
- Ethics and accountability - Takes into account how the distribution of data in training is ensured and the optimisation of methodologies and the characterisation of the VA in terms of autonomy in carrying out actions;
- Compliance - Characterises the level at which the solution is in terms of the legislation framing emerging technologies. Considers the capacity to access and inspect the technological system (audit and transparency) in order to be evaluated internally and externally.

Multidimensional readiness

The indicators

Organisational Domain

This domain seeks to assess the impact of the VA on the organisation itself, in terms of internal capacities and skills and IT resources associated with the materialisation of the VA. The indicators considered relevant for this domain were:

- Infrastructure - Characterises the level of the organisation in incorporating artificial intelligence solutions, with regard to security, interoperability, data location and availability;
- Capacities - This indicator seeks to assess an organisation's competence (maturity) to develop activities and actions in order to optimise and valorise the processes in view of the adoption of VA;
- Skills - The indicator characterises the capabilities and qualities (technical in artificial intelligence, conceptual and digital, among others) necessary and indispensable of the organisation's employees to perform tasks and actions for the operationalisation of the VA processes (implementation and management).

Multidimensional readiness

Model

Indicators
- Conversation tools
- Language processing
- Conservation methodology

Artificial Intelligence domain

It guarantees the identification of limitations, opportunities and obstacles in technological terms, comparing to the cutting lines in the progress of artificial intelligence.

Social domain

The success of an artificial intelligence solution depends on its contextualisation in terms of perception and use. This component ensures that the solution is aligned with the expectations, wishes and attitudes of the community.

Indicators
- Basic requirements
- Social performance
- Social barriers

Indicators
- Data privacy
- Ethics and accountability
- Compliance

It allows to characterise the degree of hardness and the incentives system of the current regulatory framework to support the adoption of artificial intelligence.

Legal Domain

It allows knowing the level of organisational maturity in relation to these technologies, guaranteeing the definition of organic transformation plans and the adoption of more favourable internal management measures.

Organisational Domain

Indicators
- Infrastructure
- Capacities
- Skills

Image 2 - The multidimensional readiness tool model

METHODOLOGICAL NOTE: developed by LabX and DTx, this model started from the theoretical proposal developed by the European Commission for this scope, contained in the contribution by Bruno et al (2019), plus a new dimension specifically focused on monitoring the technological consistency of the artificial intelligence solution in use and, on the other hand, ensuring its adaptation to the Portuguese context - namely in terms of its translation into measurement scales and research instruments adapted to the context.

4. CITIZENS' NEEDS, EXPECTATIONS AND “PAINS”

Citizens' needs, expectations and “pains”

Introduction

In order for this type of technology to fulfil its public service purposes, it is essential that it meets the needs and expectations of the citizens for whom this solution is intended (*Human-centered Artificial Intelligence*), as well as identifying and understanding the "pains" that citizens may feel when using it.

The aim of this exploration route was to listen to the different citizen profiles in order to rigorously monitor interactions with the VA and to make evident the behavioural layer that always accompanies the practical application of a technological solution.

Focus groups were conducted to bring together citizens, with the same profile, to systematically:

- Map citizens' usage journeys when using the VA (identifying pain points and feelings experienced during the interaction)
- Identify the paths (routes) that citizens take between the different channels made available to them to clarify their queries or search for information to deal with a given issue.

This exploration route allows us to know the relationship established between the VA and the citizens and, simultaneously, signal the difficulties experienced in its use. In this way, it is possible to identify points and aspects to improve the experience and mitigate the barriers felt by citizens during the use of the VA.

Citizens' needs, expectations and “pains”

What citizens are saying

The previous characterization and identification of the profiles should take into account the usage patterns, socio-demographic, condition of use of the service (private, professional, etc.), for example. The focus groups will reveal and identify, for each profile, the needs, expectations and "pains" in using the VA. In this format, participants have the opportunity to deal with the VA in the fulfilment of a specific task, simulating a real task of their daily life.

The model designed for these sessions was divided into five phases:

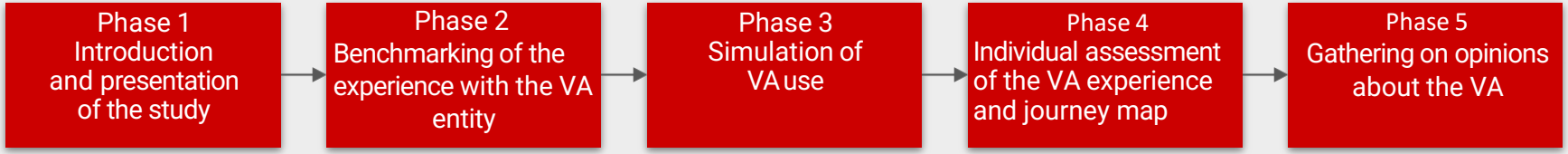


Image 3. Methodology used in the focus group sessions.

Citizens' needs, expectations and “pains”

Journey type map

The journey maps extracted from the focus groups make it possible, in a visual way, to summarise the experience of citizens during the stages involving the service provided by the VA (pre, during and post), revealing the pain points experienced and thus identifying the areas and aspects to be optimised to improve the user experience.

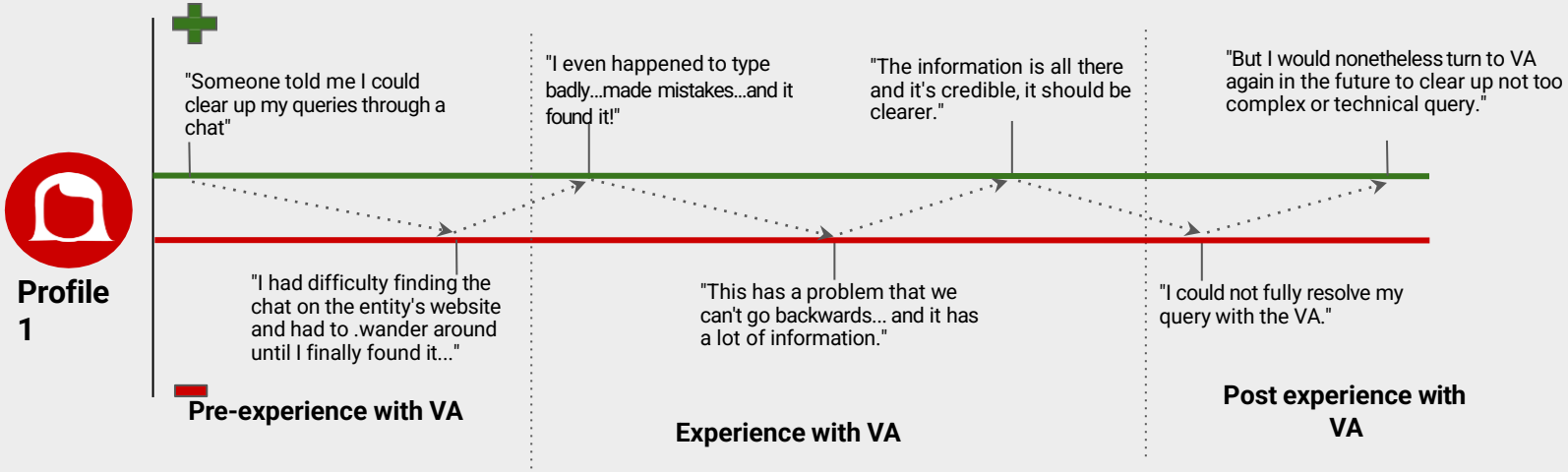


Image 4. journey type map for a given user profile

METHODOLOGICAL NOTE: The above journey map is an example of a typical experience of a given user profile, where the pain points experienced (red line, with the quotes recreated) and their comments and positive points (green line) during the service phases can be situated. The three pain points, if they were real, would have to be targeted by the entity managing the VA to improve and adapt both the interaction and the response of the VA to citizens.

Citizens' needs, expectations and “pains”

Flows (routes) in the access to the different channels

In addition to identifying and mapping the experience with the VA, it is also important to recognise its framework and connections with the various services that the entity provides to citizens. This information is obtained in "phase 2" of the methodology used for the focus group sessions, in which the experience of citizens with the entity that manages the VA and the flows between the different channels and in accessing the VA are considered, allowing:

- to know **the role of the VA in the map of the various channels and services that the entity makes available to citizens** (it allows, for example to identify whether the VA is in the first line of citizens' demand or not);
- to identify the **starting point(s) (what the citizen has already done to clarify his or her query) until arriving at the VA**.

The information provided with the identification of the routes that citizens take until they get in touch with the VA is relevant for understanding its framework and the position it occupies in the different channels and services that the entity makes available to citizens.

5.INTERACTION BETWEEN HUMANS AND MACHINES

Interaction between humans and machines

Framework

In order to evaluate the user experience of citizens interacting with the VA, a survey was made available by questionnaire, in digital format, to allow feedback from citizens regarding their satisfaction at the close of the interaction.

This instrument makes it possible to focus on three essential moments:

- **Contact with the website** (ease of navigation, availability, etc.) where the VA can be accessed;
- The **interaction with the VA**, trying to analyse how trustworthy and useful it is for citizens;
- The **interaction with the human assistant**, via chat, in case it is referred by the VA. In this way, it is possible to compare the similarities and divergences between the VA (chatbot) and the human assistant (chat).

The questionnaire allows citizens the possibility to leave their opinion, evaluating their interaction with the VA, as well as the complementary interaction with human assistants, through chat, in cases of referral, right after the conclusion of the experience.

The information gathered through this exploration route has the added value of being immediate to the interaction, because in the memory, the feelings and reactions it triggered are not attenuated by time. There is no gap between the interaction and the feedback. This way makes it possible to listen to citizens in terms of trust and empathy, clarity and objectivity of the answers and to know intentions for future use of the VA.

Interaction between humans and machines

Model of sequential conversation analysis

In order to design and assemble the satisfaction questionnaire, a model should be based on which is able to capture the 4 essential moments during the interaction with the VA; the reception, the reaction, the resolution of the issue that led the citizen to contact the entity that provides the service sought and, finally, the return. To this end, based on existing models, the following model was assembled and developed:

Conversation-driven sequence	Background	Reception	Reaction	Resolution	Return
Experience levels of the user	Sociography + Use of Services Digital	Understanding / Inference Quickness Empathy Trust Adequacy / Intelligibility Ease / Access Success / Effectiveness Return Satisfaction			

Image 5. Model of conversational sequential analysis.

METHODOLOGICAL NOTE: developed by LabX (Monteiro, 2020), this model started from the combination between a socio-linguistic proposal on the organisation of interaction sequences in human conversation (Schegloff, 2007), which allows situating the conversational analysis on the ground of everyday experience (rather than adopting an idealised model of conversation), combining with the suggestions put forward by Pia Andrews (2019) for monitoring artificial intelligence solutions adopted for interaction with humans in order to detect relevant levels of analysis in "conversational transactions" (e.g. voice-assisted technology)

6.VIRTUAL ASSISTANT METRICS

Virtual assistant metrics

The chatbot platform (backoffice)

The platform where the VA runs produces metrics that are important to consider in order to have a constant monitoring of citizens' behaviour when interacting with the VA and the performance of the VA itself. The metrics cover the following aspects:

- **Efficiency** (number of referrals to other channels when unable to respond to the request, or number of active users, for example);
- **Most popular topics**;
- **Intensity of use** (number of active users, or number of messages exchanged, for example).

This exploration route makes it possible to capture citizens' behaviours and patterns, as it enables real-time and continuous data collection. It works as a complement to the other exploration routes due to the range of metrics it enables.

Virtual assistant metrics

The chatbot platform (backoffice)

Key metrics that the VA platform must provide to tackle the three-points agenda:

Metric	Objective
No. of users (interactions) per hour and day (total users)	It allows the distribution of demand and its identification and other demand behaviours to be known.
No. of active users using the VA	Makes it possible to know if the users are recurrent in the use of the VA (if it responds to what they want) and the degree of adoption of VA (new users).
No. of new users	
Average duration of interactions;	These metrics allow to identify the level of engagement between citizens and the VA.
Total no. of messages;	
No. of average messages exchanged with users;	
No. of poorly understood messages,	It allows to identify the amount of interactions that are forwarded by the VA to the human assistant.
Goal completion rate	Captures the VA's success rate in responding to its users (interactions that were neither abandoned by users nor forwarded to the human assistant).
Most popular themes/subjects	Captures which topics are of interest to VA users.

Image 6. VA metrics table.

7.PARTNERSHIPS WITH THE SCIENTIFIC SYSTEM

Partnerships with the scientific system

Involvement of partners

For the project that supported this approach, two active partnerships were developed to technically support LabX in assessing the performance of the Tax and Customs Authority's VA applied to the methodological approach presented here: with a research centre specialised in the development of emerging technologies and, simultaneously, with a research centre able to study human behaviours towards public policy solutions.

UNIVERSIDADE CATÓLICA PORTUGUESA

Behavioral Insight Unit | Catholic University

This partner collaborated with LabX to study citizens' behaviour, looking closely at interactions and perceptions towards the VA.

dtx Digital Transformation CoLab

Digital Transformation CoLab

This partner collaborated with LabX to improve and adapt an artificial intelligence readiness benchmarking model, ensuring the shared application of field research to this AT case study.

8.KEY LESSONS

Key lessons

The application of the evaluation approach to a real case

The main general lessons learnt from applying this systemic approach to the performance evaluation of artificial intelligence solutions and their interaction with citizens, are summarised:

1. **Advantages of themes with circumscribed scopes**: VAs, in early stages of entry into service, approach optimal performance the more circumscribed their scope is. Targeting a theme allows responses to be more stable and more satisfactory;

2. **Priority on "neutral" topics**: in the initial phase of introduction in public administration, VAs benefit from avoiding topics with high complexity (with many exceptions or context-dependencies) or topics that require answers that require guarantees and could, for example, serve as evidence. This is not what citizens expect for early stages of VA solution maturity;

3. Function as a **first line of triage**: this service solution should always be seen in articulation with the system of alternatives, the other service channels that the entity makes available to citizens, and may take the opportunity to occupy a distinct and complementary situation before these options from its assumption as a first line of triage, thus allowing the pressure on other channels (face-to-face, telephone contact, etc.) to be mitigated.

4. **VA's maturity and experimental approach**: to improve the VA's understanding of requests, quality of responses and problem solving, it is important to continue its development process and evolve its maturity, which unequivocally calls for the adoption of iterative experimental development cycles.

Key lessons

The application of the evaluation approach to a real case

5. Maintaining the **interactional nature through conversational services**: the choices made for the design of the VA as a "conversation partner" to clarify queries is in line with the progress registered in the applications of this technology at international level;

6. **Readiness levels**: the identification of readiness levels makes it possible to situate the level of development and maturity of the VA and establish a path of evolution, through a concrete plan of improvements, always associated with the results obtained by the other exploration routes;

7. **Sharing and collaboration within the Public Administration**: the development of this approach to evaluate the performance of VAs adopted by the Public Administration from the embryonic phase, the articulation between similar initiatives within the Public Administration, ensuring transversal parameters for the digital infrastructure, the integration of systems, and the accumulation of data and learning, among others;

8. To persist in an **experimental logic with iterative development cycles**: to guarantee the centrality of these solutions around citizens' needs (since these are in constant mutation) implies keeping a permanent follow-up in terms of testing, monitoring, evaluation; on the other hand, given its expected relevance in the future, a constant investment must be kept, even if through controlled experiments, allowing an incremental evolution of the solutions;

Key lessons

The application of the evaluation approach to a real case

9. The methodology, even in an experimental version, showed potential to be used by the Public Administration. The methodology tested with a real case, proved to be a powerful instrument to be scaled and used by public entities that want to assess the performance of their VA and subsequently optimise the VA. The application of this methodology to the VA of the AT's Finance portal contributed to: (i) obtain a current diagnosis of its VA; (ii) identify a set of recommendations to evolve its VA and increase its performance and adequacy in responding to citizens, and meeting their needs and expectations;

10. **Citizens are willing to use and resort to this type of solutions,** even if the interaction does not resemble the characteristics of an interaction with a human (human-like interaction), they understand the potential of these solutions for the present and see enormous potential for improvement to better meet their needs in the near future;

11. The issues of **transparency and credibility are seen by citizens as key aspects** to boost their adoption;

12. **VAs require human supervision and constant training.** It is a common misconception that AI systems (using machine learning or hybrid systems, for example) learn alone and operate completely on their own, without human support. Training the VA with a view to improving and adapting its responses requires the team that manages it to do a demanding job of analysing the conversations to ascertain errors and carry out constant training of the VA.

CENTRO PARA A INOVAÇÃO NO SETOR PÚBLICO

ama AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA

Citizen-centred artificial intelligence

An approach to virtual assistant evaluation in public services

labx@labx.gov.pt | linkedin.com/company/labx-govpt/
facebook.com/labx.gov.pt | @labx_govpt