Deloitte.

Publication Office of the European Union

C.3: Technical documentation

Advisory & Consulting, Digital & Technology

February 2021

Table of Contents

Revision history

TABLE 1: REVISION HISTORY		
Comments	**Section**	**Deliverable**
"certain documents are retrieved that have empty metadata fields" – how do you conclude on the origin on the issue as the mitigation is metadata enrichment?	2.4.2 Risks and production challenges	C.2
What does it mean monitoring of the bot/assistant? How can we do it? How can we ensure and measure it remains all the time relevant? How can we measure it remains responsive?	3. Industrialization approach (Stream C)	C.2
How can we simple measure and monitor?	3. Industrialization approach (Stream C)	C.2
How to handle the control of the quality under the MSP model? (as an example today the search functionality: we have the golden queries and an index of relevancy measured monthly for each and every golden query. The config of the search we don't care much but the output)	3. Industrialization approach (Stream B) *This will be assessed during Stream B of implementation as part of the Operational process review.*	C.2
What would be the Service Level Requirements for the assistant under MSP model as compared for instance to the ones today in Search	3. Industrialization approach (Stream B) *This will be assessed during Stream B of implementation as part of the Operational process review.*	C.2
In order to start an industrialization project we will need requirements and a price estimation. The report does not give any indication of what would be requirements and how we can come to a comprehensive list of requirements in order to have a price estimation to start implementation.	2.4 The way forward to OPIA industrialization, 3. Industrialization approach *No pricing included in this document*	C.2
Furthermore, from methodology perspective, we run OP Portal under MSP FWC. We expected a specific approach on bot/Assistant industrialization considering our specificities under MSP as you know that we are running under MSP. For example points like DIS.6, TES.1, DES.3, DEP.2, S.1, S.2, S.5 are of very little interest for us under MSP, while others like TES.2 and C.3 are crucial and needs probably heavy preparation, but how can we prepare and especially what should we prepare? (e.g simplistically, how would we test the assistant in Lithuanian and make sure is relevant? How	3. Industrialization approach	C.2

could we organize the UATs, what should we look at rigorously?		
The methodology described is quite generic applicable to any project whatsoever	3. Industrialization approach	C.2
In this report, we would expect a roadmap with key activities, milestones and a sequence timeline (approximatively) which would take. How long would take us if we order it for 1 language, for 5 languages for 24 languages?	4. Overall planning, 2.4.3 Additional features to be considered	C.2
Teaching/ is based on three use cases here. What will be the case for production? How can we imagine the teaching process?	3. Industrialisation approach (Stream B)	C.2
Training LUIS – if nothing is provided out of the box for intents like those in chapter 6.2 (Foreign, Greetings, Mood, Name, etc)…how should we tackle this given the 24 languages?	2.4.3 Additional features to be considered, 5.1 LUIS prebuilt packages	C.2, C.3
Applying NLP and ML techniques, the OPIA Intelligent Assistant is able to recognize relevant Entity values. Some entities we have (Authority tables, taxonomies while many he needs to learn from user queries. As this is the core of the assistant "understanding" to what degree the NLP and the ML is applied to the other languages than EN and FR? Could we classify languages like "excellent support", medium support and low support for NLP and ML available today?	2.4.3 Additional features to be considered, 5.1 LUIS prebuilt packages	C.2, C.3
What aspects should we consider in the design (choice of bot per language, bot language independent, speech per language, etc)	2.4 The way forward to OPIA industrialization	C.2
What are the top priorities we should prepare in our team? For example, top 3 or 5 Skills, experts and expertize?	3. Industrialization approach (Stream A)	C.2
Are there any elements, which require very specific Azure subscription? (e.g. cognitive services)	All the components currently foreseen, can be used under basic Azure subscription.	
What are the top prerequisites before starting the industrialization project?	2.4 The way forward to OPIA industrialization	C.2
What are the evolutions of search? Is search the only service the BOT should rely on?	5.1 LUIS prebuilt packages	C.3
What are the main top three security challenges and top 3 vulnerabilities associated with such a bot/assistant? More particularly what are the top three vulnerabilities of the client facing interface/API service (is it the bot framework?) aka the Endpoint?	2.4 The way forward to OPIA industrialization, 2.1 Microsoft Azure	C.2, C.3
XMl vs JSON In your recommendations (3.2.1) you propose two mitigations methods but both consider translating XML to JSON while none	2.4.1 Proposed backlog from OPIA POC	C.2

refers to potentially having the OP Portal Search API answering with JSON instead XML. Would this not be natively more performant?		
All the services are under one tenant in Azure? Can the perimeter of the services used be regrouped under one domain name? (e.g. https://op.europa.eu)	5.6 Configurations	C.3
Are there all services used described? In the technical workshop it looked like more apps/services described. A diagram like the one in the technical workshop summarizing would be also welcome.	5.6 Configurations	C.3
6.2: Search intent: why only one intent for these categories? Could we explain the choice and if this is something, it should/could be different for industrialization.	5.1 LUIS prebuilt packages	C.3
"At the moment the speech and written assistant is separated in two bots." (two or four?)Such detection is possible through middleware, which is a solution that will be investigated for the deployment of the OPIA. – could we have this conclusion a bit more explicit, the choice made and explanation clearer and refined?	2.4.1 Proposed backlog from OPIA POC, 4. Conclusion	C.2, C.3
6.2: List of all Intents and trained Utterances: which ones are there out of the box and which ones had to be "inserted (and how)" by you? The impact per language? It worth be specified in the report	5.1 LUIS prebuilt packages, 2.4.3 Additional features to be considered	C.3, C.2
Out of all services is (are) the Endpoint? Could you maybe make it clearer also in the chapter 2.1. What (the name) is the client facing service (or api)? (the "thing" to which a user or external app can connect to interact with the assistant?	2.1 Microsoft Azure	C.3
Crucial: There is no reference to how can we demonstrate the added value of this bot/assistant. How could we drive it now from a POC to a positive decision to go for industrialization? SWAT analysis, pro and con, basic steps of communication and outcome expected? What are the top 3 or 5 arguments to go into an industrialization? Why would we do it? (I guess not just because it's fancy). How much (approx) will cost for 1 or 24 languages?	2.4.4 Rational behind implementation	C.3
How helpful was Microsoft and what have you used from their input specifically?	Microsoft was available for advice on their products and assistance with certain challenges in development. We had regular touchpoints and the Ms team could inform us about version incompatibility between Bot Framework Composer	

	and the latest version of Adaptive cards, which helped to move with the timeline. Microsoft was helpful to suggest alternative products that could be alternative solutions for the EuroVoc mapping. Their input was specifically used in the initiation phases to ensure the relevant selection of components for the POC.
What do you think we could improve and focus on?	The focus should be on having an assistant that is user-friendly and achieves the goal of helping users search more effectively. UAT testing should ensure usage is user friendly and comprehensive. To be more accessible to all users, an addition would be to include German and Spanish as languages and by doing so covering the four most frequently used languages on the Portal. This can be done following the same approach or making use of translation services to accelerate implementation.
Any remark for OP Portal improvements through the context of the OPIA experience you had?	By working with the OP portal, two main points came across: • Changing the OP API response format from XML to Json would allow a direct communication for the bot. This would be a less complex architecture and limit additional costs and components. • We have noticed that the metadata used by the bot is not always available for all documents on the portal. As the OPIA is highly reliant on metadata to guide users, looking into a metadata enrichment strategy would improve the search experience and enable the discovery of more documents.

Glossary of terms

TABLE 2: GLOSSARY OF TERMS	
Technical terms	**Description**
AI – Artificial Intelligence	Computer technology that allows machines work in an intelligent way.
API – Application Programming Interface	A software intermediary that allows applications to retrieve information from one another.
Entity	Entities are important pieces of information that can be extracted from the user input that is relevant to the user's purpose. These piece of information will be identified and stored to extract exactly what the user needs. An example of an entity can be the author of a specific article.
Intent	Intent refers to the goal the customer has in mind when typing in a question or comment. An Intent represents an idea or a concept that can be contained within a message sent by the user, called an Utterance. An example of an intent is the fact that the user wants to search for a specific topic.
LUIS – Language Understanding	Cloud-based conversational AI service that applies custom machine-learning intelligence to a user's conversational, natural language text to predict overall meaning, and pull out relevant, detailed information
NLP – Natural Language Processing	Subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language.
POC – Proof of Concept	A realization of a certain method or idea in order to demonstrate its feasibility.
RDDA – Reading Disability and Document Access	Project mandated by the European Parliament, which aims to develop and make available appropriate instruments to ensure that people with reading disabilities have equal access to the documents or publications and websites of EU institutions.
Utterance	Input from the user that can be any message typed in a conversation to the assistant. From this message the key Intents will be recognized.

1. Introduction

The OPIA Intelligent Assistant is developed in line with the EU Parliament mandate of the "Reading Disability and Document Access" (RDDA) Pilot project. The intelligent assistant's goal is to increase usability of the search portal. Two main functionalities are in scope for this POC: speech-to-text and the intelligent search assistant.

A complete technical description of the system will be provided in this document. This covers:

- The description of the technologies used
- The API configuration
- The data science approach steps to cover the creation of the intelligent virtual assistant
- The training of this virtual assistant will be detailed
    - The whole flow process to achieve the Proof of Concept (POC)
    - All documentation regarding Intents, Entities and other training data used for the OPIA cognitive model training.

The goal of this deliverable is to cover and detail the technical aspects of the OPIA POC.

2. Description of technologies used

As part of the vendor selection process, the top three different technologies have been was evaluated based on benchmarking criteria that covered the solution profile, functional and non-functional requirements (see Deliverable A.2 Vendor selection). After weighting the selection criteria, Microsoft Azure has been chosen by the Publication Office to build OP intelligent assistant POC.

2.1 Microsoft Azure

The vendor selected by the Publication office for the implementation of the OPIA POC is Microsoft Azure that offers a wide variety of services within conversational AI. The main components used during the POC includes:

- **Language Understanding (LUIS)**: This is a cloud-based conversational AI service that applies custom machine-learning intelligence on which cognitive services can be integrated such as the Azure speech service that converts spoken language requests into text. LUIS adds prebuilt domain models or custom model to determine the Intent of the user.
- **Bot Framework Composer:** Framework for the creation of an intelligent assistant using Azure cognitive services. Bot Framework Composer integrates LUIS for sophisticated composition of bot replies using language generation. With this framework, it is possible to use visual editing canvas for the conversation flows and adapt the Entity and Intent recognition.
- **Logic apps:** Cloud service to automate a workflow in an application within Azure. It is possible to use a combination of built-in connectors to convert XML files to JSON for each HTTP request for example.

The OPIA is developed using Ms Azure's Bot Framework Composer by defining the different user Intents and the conversational flow. The Machine Learning model from LUIS uses Natural Language Processing (NLP) techniques to cut through the complexity of unstructured data and generate insights that can feed the search on the OP Portal. The raw messaging interface where the search-assistant-end-user-interaction occurs is a middle layer that parses the text and derives insights. The OP Portal search API is called to perform the intended action in order to receive the relevant documents and metadata, where Logic Apps

performs the conversion to share enable the OPIA Intelligent Assistant to display the information to the user (see 3. Developing the OPIA Intelligent Assistant).

As a final step the OPIA is deployed on the Azure portal, via an endpoint that is to be accessed by the Publication Office website. This takes place on the Azure Portal. Specifically, the endpoint appears on the Bot Channel Registration (where all different services and features are gathered) and through the Azure Web App (location where the OPIA is stored and processed). When the OPIA POC is deployed, the end user can interact with the OPIA from the OP portal through a direct line channel to the Azure portal where the endpoint is configured.

2.2 Publication Office search API connection

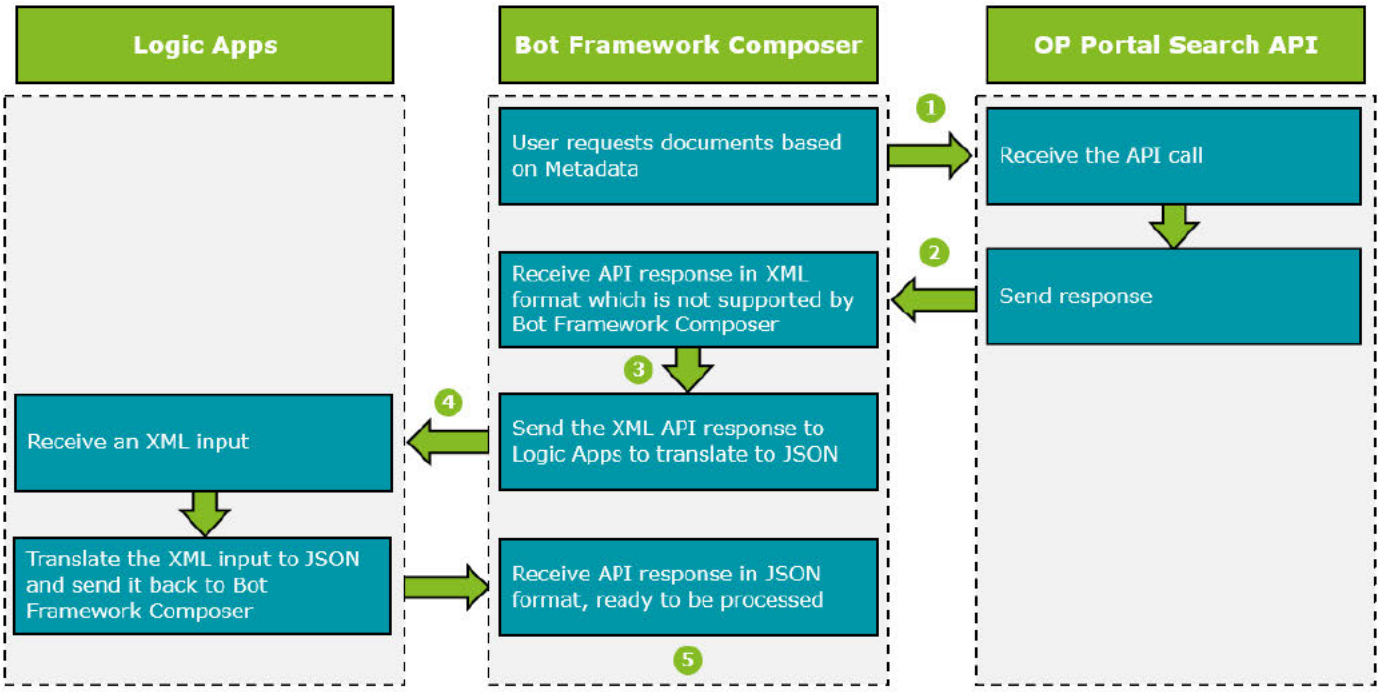
The Publication Office search API provides multiple methods to search and display relevant documents based on the user queries sent through the virtual assistant. The functioning of the virtual assistant relies on calling the right function with the right known Entities. There are multiple useful API functions, such as 'browsing by subject', which returns all documents matching a certain subject or 'find similar' which allows to retrieve a list of similar documents based on the document identifier given as a parameter. For 'simple' search, the API filters for one metadata and for 'advanced' search for more. Other options such as language, subject and author is also included in the possibilities. To use 'advanced' search, several related API functions must be called to fit with the metadata provided by the user. More detailed functions are described in the official API documentation.

> *Example: If a document search Intent is recognized with the Entity "tax regulation" as theme, the virtual assistant can perform a 'simple' method search where it only displays relevant documents for the theme: "tax regulation". However, it can also perform an 'advanced' search to fill multiple fields. The user provides "I want tax regulation for the year 2000", then the search API should detect the theme: "tax regulation" and the publication year: "2000".*

Each API request requires a basic authentication to identify the license responsible for the request. The process followed can be described in the following steps:

1. The API call with the user's input starts with the HTTP POST request to the Publication Office Search API
2. The response is received in XML format - which is not directly interpretable by the virtual assistant. The Intelligent Assistant sends a POST request to an Azure Logic App to convert the XML response to a JSON format (the format expected by Azure Bot Framework Composer)
3. This metadata received from the Publication Office Search API is stored in different variables, including the number of documents filtered, the authors, documents titles, subjects, period etc. These variables are used to provide the final results to the user.
4. The process repeats for each response from the user that is detected by the OPIA Intelligent Assistant as a relevant entity to make a post request to the Publication Office Search API followed by a call to the Azure Logic App for a XML to JSON conversion.
5. At the end of the conversation with the user, the virtual assistant will display the two most relevant documents and give a quick overview of the search results to the user. The user can select "find more" to retrieve additional relevant documents based on the query given as a parameter. It is possible to prune user's search by storing new entities and calling additional API requests.

FIGURE 1: API CALL PROCESS

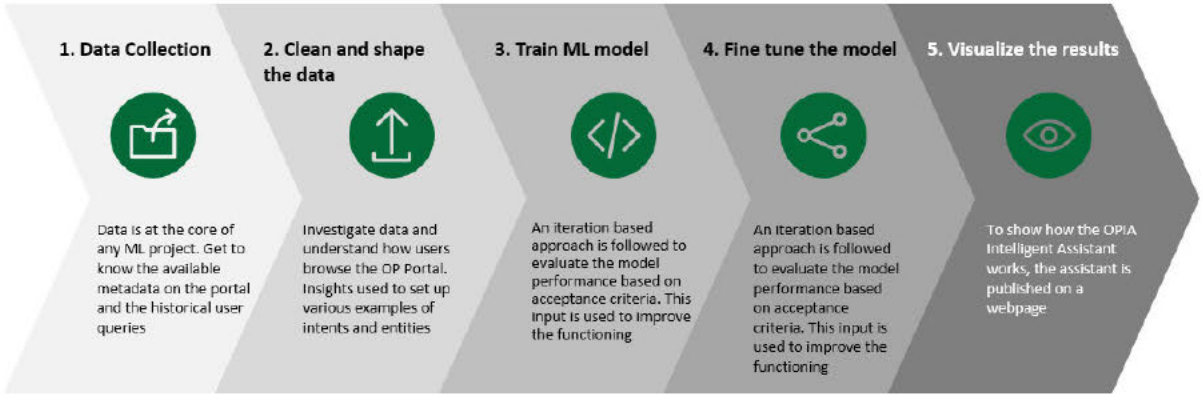


By using the search API, the information provided by the user to the OPIA Intelligent Assistant can be effectively filtered using LUIS Cognitive Services to send requests to the Publication Office Search API. The process is followed and the OPIA Intelligent Assistant returns a selection of document to the user. With the environment selected and the connection to the API set up, the development environment for the search assistant is ready.

3. Developing the OPIA Intelligent Assistant

The development of the OPIA Intelligent Assistant was completed based on the five steps of the data science approach. The intelligent assistant's goal to increase usability of the search portal is kept in mind throughout the sprints. Two main functionalities speech-to-text and the intelligent search assistant functioning with text input is to be developed during this POC.

FIGURE 2: THE MAIN STEPS TO DEVELOP A VIRTUAL ASSISTANT



3.1 Data Collection

The scope of the OPIA includes documents under the categories EU Law and EU Publication. All the documents have been indexed with metadata and the content is available in 24 languages, of which English and French is in scope for the OPIA POC. To make use of the available data, the portal metadata and historical search analytics is used.

The OP Portal Search API is available to access the data and all relevant metadata of the OP portal. Table 2 lists the metadata on the OP portal and provides a description of each.

TABLE 3: METADATA

Metadata	Description
Title	Title of the document
Type	Type of document (law, report, survey, …)
Author	Name of the author(s) of the document
Publication date	Date when the document has been published
Subject	Key topics describing the document
Identifier	Unique ID of the document
Language	Language that is used in the document
Format	Format of the document (html, print, pdf, …)
Thumbnail	A miniature replica of the document's front page
Private author	Hidden authors of the document

Additional available data was collected from historical searches made through the OP Search portal. Actual search data provides useful analytics of user searching habits that can inform the training and structure of the search assistant to assist users in their search. 282 968 queries (531 queries every day on average) on the OP Portal between 1st April 2019 and 15th September 2020 indicate the following:

FIGURE 3: USERS QUERIES ANALYSIS

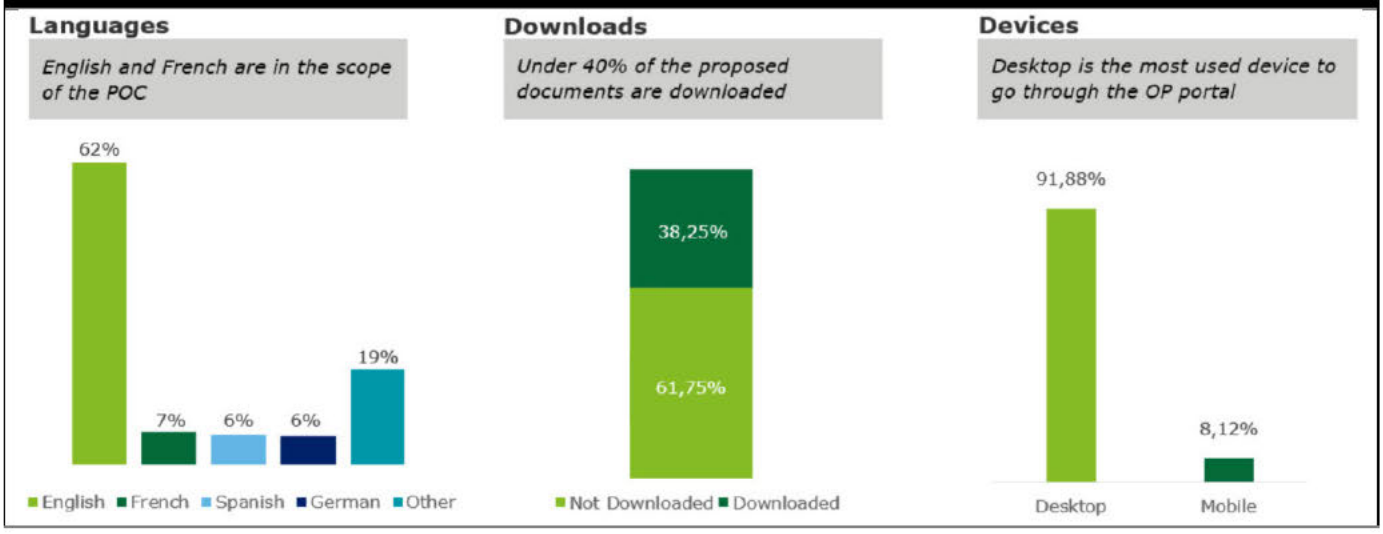


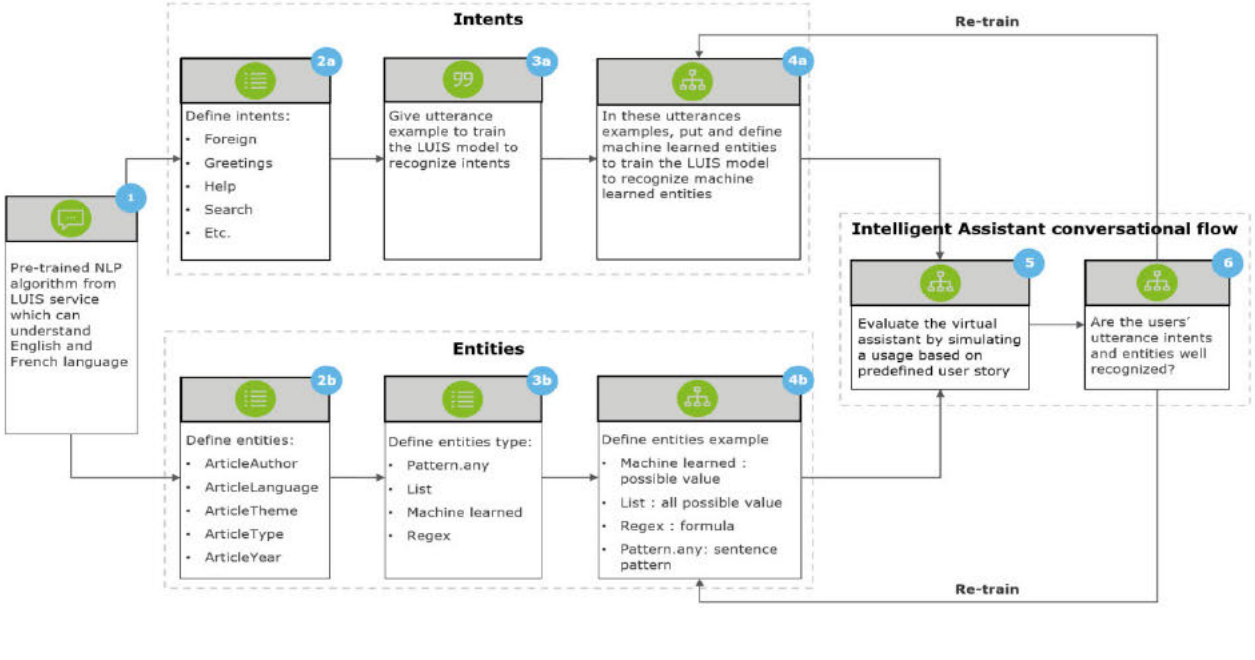
Table 3 below summarizes the attributes available in the historical search analytics.

Table 4: Attributes	
Attributes	**Description**
Time	The date when the file was created
Query text	Describing the category of the document
CellarID	The unique ID of the document
URL	The link to the file
Title	The header in the document
Type	The support of the document: mobile/desktop
Page views	The page of the document viewed
Visits	Number of people who visit the document
View time	Amount of time the document has been viewed
Downloads	Number of download of the document
Subjects	Subjects of the document
Authors	All the authors from the document
Language	Language of the document

3.2 Clean and shape the data

Investigating and understanding the data can provide valuable insights to the design and conversational flow of the OPIA. These insights are used to set up various examples of Intents and Entities. The 3.2 Clean and shape the data below describes the main steps to build this intelligent virtual assistant.

Figure 4: The main steps to train an intelligent virtual assistant



In order to train the virtual assistant, certain training data needs to be clearly defined. This includes Intents and Entities that are used to improve and customize the OPIA Intelligent Assistant's understanding of certain concepts and texts. Intents and Entities are relate to the overall conversational flow of the search assistant and certain search triggers should be assigned upfront to guide the different possible user paths in a given conversation.

The first step of a NLP algorithm is to define user Intents. An intent represents an idea or a concept that can be contained within a message sent by the user, called an utterance. A well-trained NLP algorithm can understand several intents in a discussion. To do this multiple examples of Intents should be provided to train the virtual assistant effectively. Recognizing Intents allows the virtual assistant to understand the meaning of users' messages and then extract relevant variables. These relevant variables are called Entities and are crucial data to prune our search through the Publication Office Search API. An Entity is any word or series of words that consistently refers to the same thing in a discussion. To ensure a large coverage, Entities are defined based on the OP Portal metadata. Similarly to Intents, the virtual assistant also needs many examples of Entities for training purposes.

Here is a non-exhaustive list of the Intents as well as some examples. The fill list is available in the appendix.

TABLE 5: INTENT DEFINITIONS

Intent	Description	English	French
Foreign	The Intent picks up different languages when someone is starting a conversation with the bot	Asalaam alaikum Ciao Cześć	¿Qué tal? Asalaam alaikum Ciao
Greetings	When the user greets the bot the user receives a greeting back from it	¿Qué tal? Asalaam alaikum Ciao	Bienvenue Bjr Bonjour
Help	This Intent aims to provide assistance to the user while asking for additional help using the search assistant	Can I get help? Can you assist me? Can you give me a hand?	J'ai besoin de ton aide J'ai besoin de ton soutien Je me demande si tu peux m' aider
Mood	The bot can interact with the user when asked how she/he is feeling	Are you fine? Are you ok? Are you well	Comment ça va ? Tu vas bien ? Comment tu te sens ?
Name	When the user asks for the bot's name, she/he receives an answer from it	Do you have a name? How can I call you? How do people call you?	C'est quoi ton nom ? C'est quoi ton prénom ? Comment devrais- je t'appeler ?
Questions	This Intent picks up if the user asks general questions	Do you have a family? Do you like movies?	Quel est le temps ? Quel jour sommes-nous ?
Search	These are the main search Intents and provide numerous ways how the user can search a specific document	Search Search an article about right of citizen Search for a doc with info about renewable Technologies	Cherche-moi un document qui met en avant l'inégalité des sexes Cherche- moi un document qui met en avant protection de l'environnement

Table 6 shows a non-exhaustive list of Entities. The full list is available in the Appendix.

TABLE 6: ENTITIES FOR THE OPIA PROJECT		
Entities	**Examples in English**	**Examples in French**
Article author	• European University Institute • European Data Protection Supervisor	• Direction générale de la mobilité et des transports • Commission européenne
Article language	• French • English	• Francais • Anglais
Article theme	• Data privacy regulations • Energy efficiency	• Réglementations environnementales • Collection de données
Article type	• Html • Pdf	• Html • Pdf
Article year	• 1970 • 2013	• 1970 • 2013

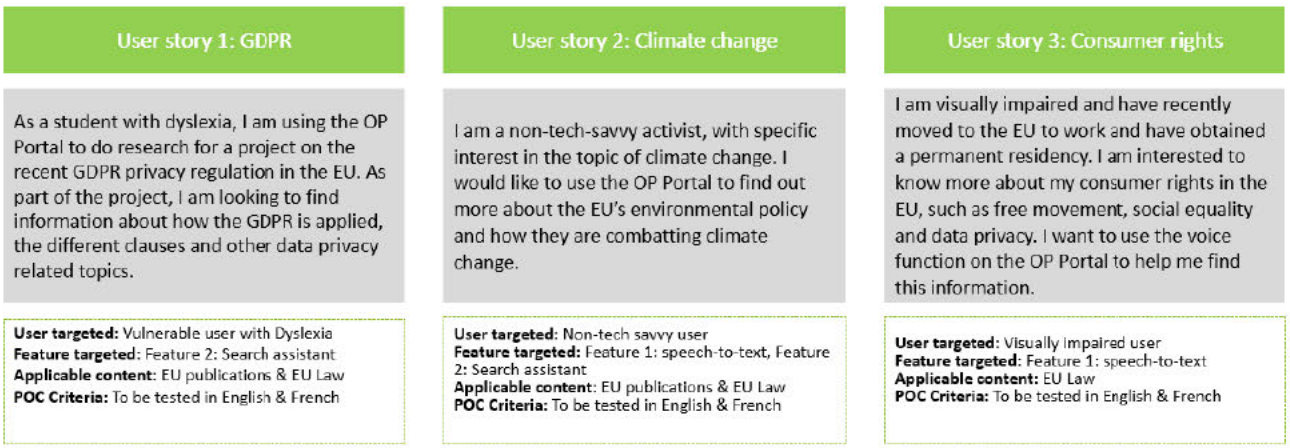
When the user sends a query to the virtual assistant, if an Entity such as the theme or the publication date of the article is detected with LUIS recognizer, it is stored as a value and the document search will be filtered with these values. To train the model to recognize specific Entity, their potential locations were defined in multiple queries using patterns to cover multiple forms for the same expression. The virtual assistant also uses advanced features like the adaptive expressions and conditional responses to adjust its answers to the user's entries.

Once an Intent is recognized, the bot will enter a specific flow to converse to the user. For example if the "Greeting" Intent is recognised with the query "Hi", the assistant might reply with "How can I help you today" for example. If the "Search" Intent is recognized, the virtual assistant will guide the user in the research a relevant document. However, like every machine learning algorithm, the Intents and Entities detection will perform based on training and a great deal of attention should be added to the phrases used and the conversational flow set. The algorithm also interacts with the document metadata on the portal making it dependent on completed and correct metadata.

3.3 Train ML model

The cognitive LUIS model is based on NLP. The assistant's goal is to predict unseen Intents and Entities in discussions based on what it learned during its training. Three predefined user stories form the basis of the training of the machine learning algorithm for the OPIA POC.

FIGURE 5: USERS STORIES



The LUIS machine learning model is a cloud-based conversational AI service that applies custom machine-learning intelligence to a user's text or speech input. This is used to predict the overall meaning and extract relevant information that the assistant uses to query the portal. There are different ways to train different Entity types in LUIS. The following four types are relevant

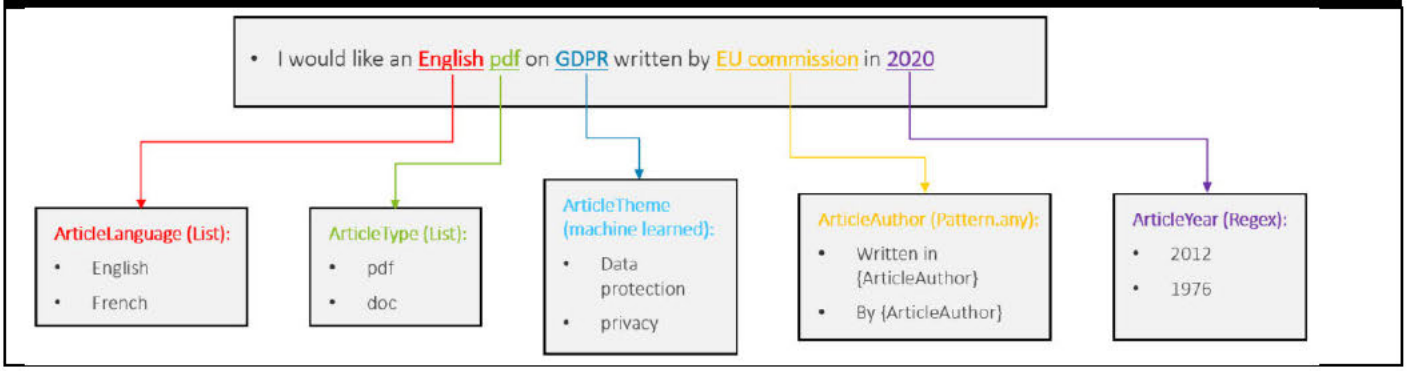
TABLE 7: ENTITIES TYPE DESCRIPTION

Entity type	Description
Pattern	This is s a variable-length placeholder used only in a pattern's template Utterance to mark where the Entity begins and ends. For Entities, where full lists cannot be defined, patterns such as the possible location of an Entity in a sentence can help the Smart assistant detect them. This is applied in the OPIA Intelligent Assistant to detect Authors.
Machine learned Entities	All Entities defined are detected automatically in training Utterances. By defining many possible values, gather by users stories theme, and by training the model to detect them with many Utterances, the LUIS model will be able to detect Entities that it has never seen before. It suits perfectly for theme Entity which is not a finite list and allow grammatical errors.
Regex	Regular expression which detects specific combination of number. It is specifically useful to detect year in any sentence.
List	Possible values for a given Entity, it requires to match exactly one of the predefined Entities examples in the list or one of it synonyms also predefined in the list. These Entities are easily predictable and don't require machine learning to be detected. For instance, it suits perfectly for document type where there are a known list of possible document type on OP Portal.

The machine learned Entities type (see Table 6) requires examples of Intents and Entities to be able to learn and generalize on new user input, as it will never be possible to provide a list of every possible query the user will search. Additionally, this will also account for misspelling, grammatical errors or a different tone used by different users. The machine learned Entities will allow the virtual assistant to detect something more meaningful than a simple list of possible values. To train the model many possible user Utterance should be provided that contains relevant Entity values. By providing these examples, the ML model applies NLP and can learn to generalize on new user Utterances.

Example: One of the training Utterances is: "please help me to find a document about {@ArticleTheme=freedom of speech}". The OPIA Intelligent Assistant learns from many examples and different sentence formulations in order to be able to detect different theme Entities in future Utterances from real users. Let's say the user uses the trained search assistant and writes "I would like a document on climate change". The OPIA Intelligent Assistant would be able to detect {@ArticleTheme= climate change}.

FIGURE 6: HOW THE VIRTUAL ASSISTANT UNDERSTAND USER UTTERANCE?



Applying NLP and ML techniques, the OPIA Intelligent Assistant is able to recognize relevant Entity values that doesn't exactly match the example lists. This means that many misspelling and grammatical errors could also be supported to a certain extent. Additionally, patterns can be defined to assist the learning process. For instance a pattern can be "look for", "search for" and "try to find". This could be viewed as providing the model with synonyms that allows to detect the right Intent and avoids to define each Utterance separately. The ML model detection will never be perfect and the flow, options and training patterns of the virtual assistant should be carefully designed. One key feature of a well-trained ML model is that it can easily be expanded to include a new scope to match future requirements.

When a given input matches multiple dialog nodes, the intelligent assistant shares a list of the most relevant options and asks the end-user to select. The user is also able to click to see more options, filter and refine current search or continue to the advanced search on the OP portal. The assistant can also redirect the user to contact the help desk. In the case where the user Utterance is not understood, the virtual assistant will redirect the user or ask the end-users to clarify their meaning. The following search guidance options are included functions:

- Prompt questions to direct user's search and allow filtering
- Include interactive tick boxes and buttons for selection
- Use additional metadata to prune the search
- Have speech-to-text functionality

Furthermore, LUIS provides many test features such as the confidence score evaluation of expressions and Entities, to highlight the strength of the virtual assistant and its weaknesses. LUIS model can always be improved regarding what the users have asked and correct errors. LUIS can facilitate storing and syncing of reference data. It allows direct links to OP metadata and updates through Azure Logic Apps. Hence, the LUIS model will be sync to be up to date with the last backend changes and therefor remain relevant.

Language inclusion must be considered when training the OPIA Intelligent Assistant. Both English and French is in scope for the POC. To teach the virtual assistant to understand multiple languages, there are two possible approaches. The first approach is to use a translator to adapt the model to different

languages, with this approach can optimize development time and processing resources, with the trade-off that the virtual assistant is processing translated content and thus may not always respond properly to the user. The second approach, which is applied in the OPIA POC, is to define a separate model for each language (English and French for the OPIA). This requires additional input, but leads to a better understanding of the user Intents and is a more accurate method. The virtual assistance response will be as intended and easy for the user to understand/ follow as no translation is applied.

3.4 Fine-tune the model

A sprint based training approach has been followed to evaluate the model performance based on acceptance criteria. Three iterations, each following one user stories will simulate real cases. These iterations will result in retraining the virtual assistant, redefining Intents and Entities, maintenance of the model and the dialog flow to fit with users' needs and different ways to interact with the virtual assistant. The iteration-based evaluation of the model performance based on acceptance criteria is used as input to improve the functioning of the OPIA POC.

By scoring and evaluating the virtual assistant at each iteration, it would be possible to identify the current backlog and assign possible improvements to ensure an OPIA Intelligent Assistant aligned to the POC specifications. The OPIA's POC will be evaluated based on three pre-defined user stories validated by the Publications Office:

- **GDPR**: As a student with dyslexia, I am using the OP Portal to do research for a project on the recent GDPR privacy regulation in the EU.
- **Consumer rights:** A visually impaired individual wants to use the voice function of the OP Portal to get some information about his consumer rights as a permanent resident in the UE
- **Climate change:** A non-tech-savvy activist having specific interest in the topic of climate change wants to use the OP Portal to find out more about the EU's environmental policy.

Acceptance criteria covers nine aspect of the virtual assistant and define the functionalities that the search assistant should cover as described in the table below (see deliverable OPIA_workshop_2_POC_Delivery). This would ensure a satisfying user experience, more guided search, smooth interaction and the provision of relevant information or alternative option to continue using the OP portal.

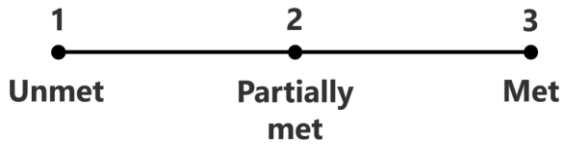
TABLE 8: ACCEPTANCE CRITERIA

Code	Name	Description
AC.1	Search guidance	Does the intelligent assistant help the user formulate search? Is the user guided to search better than without the OPIA? Does the assistant facilitate usage of OP portal metadata? Does the assistant efficiently direct the user to the needed search?
AC.2	User Experience	How is the overall experience (based on user profile)? Does the user easily find the search assistant? Is the assistant accessible based on design best practices? Does the assistant always answer?
AC.3	Response speed	Does the search assistant give the user enough time to read and grasp the information shared? Is the response speed to provide a document suitable?
AC.4	Conversational robustness	Is the conversation flow clear? Are all buttons functional? Did the assistant properly redirect or clarify the situation when the user was not understood?
AC.5	Interaction	At what point of the journey is the user discovering the assistant? Can the user exist the assistant but also activate it again at any point of the journey? How easy is it to navigate through the conversation history and direction?

AC.6	Speech-to-text functionality	How well was the user understood when using the speech-to-text functionality? Was the search assistant able to adapt to different languages (English & French)?
AC.7	Search efficiency	How many questions does the intelligent assistant ask before suggesting a search result to the user? Is the assistant able to help the user by asking only a few questions?
AC.8	Search conversion	Does the user complete the search when initiating the assistant search option? Does the use of the OPIA provide narrow suggested search? Does the OPIA propose links to the user or additionally suggest a search launch?
AC.9	Relevance of documents	Are the proposed search links relevant to the topic the user discussed? Does the first page of returned searches contain information needed by the user? How relevant are these results?

The scoring scale for evaluation ranges from 1 to 3, with 1 being the lowest score (i.e. unmet criteria) and 3 the maximum score (i.e. met criteria). A detailed description of the expectations at each level of the scope is included in Appendix 6.5.

Evaluation scale



3.5 Visualize and communicate results

In line with the POC planning, the user stories has been introduced one by one in the sprints. User story 3 was included first, ensuring that the OPIA could understand topics related to climate change. User story 1 was incorporated thereafter, looking into vocabulary related to privacy rights and GDPR. Lastly, story 2 was included in sprint 3 to cover consumer rights and related queries. The virtual assistant was evaluated within 3 sprint evaluations based on 9 criteria that describes the robustness, the relevance of the flow, the user experience and other related points. Sprint 4 to 6 was aimed at improving the remaining backlogs from the first sprints covering all user stories and reviewing both English and French language flow and model language understanding.

> *Example: If the voice feature is not included into the virtual assistant, then the score for acceptance criteria 6: Speech-to-text functionality will be "Unmet". Once the voice functionality for French and English voice has been included into the virtual assistant but with remaining blocking points, then the acceptance criteria 6 score was "Partially Met". Finally, when the speech feature is successfully implemented and configured, acceptance criteria 6 score will be "Met".*

Table 8 gives an overview of the sprint cycles, indicating how the different languages had been included as well as different features for testing. Throughout the progress, different testing environments have been used and finally the bot had been tested on a static website view of the portal to also be able to test the size of the chat box and the font that will be displayed to the user.

TABLE 9: SPRINT LOG				
Date	**User Story**	**Functionality**	**OPIA test location**	**Key actions logged**
Sprint 1 (11/11)	3 - Climate change	Written (EN)	Bot Framework Emulator	• Adapt search flow • Add metadata filtering • Review LUIS detection model
Sprint4 (18/11)	1 - GDPR (Backlog: US 3)	Written (EN & FR) Speech (EN)	Bot Channel Registration Windows Voice Assistant Client (Speech)	• Implement adaptive cards • Include French (written) • Add different dialog flows • Work around dead ends • Speech added in English
Sprint 3 (25/11)	2 - Consumer rights (Backlog: US1, US3)	Written (EN& FR) Speech (EN)	Storage Account on Azure	• Cognitive model trained • Dialog flow has no dead ends • Deploy virtual assistant to static website on Azure
Sprint 4 (26/11)	All user stories	Written (EN& FR) Speech (EN & FR)	OP portal static website	• Speech in both languages • French LUIS model improved • French dialog flow reviewed
Sprint 5 (03/12)	All user stories	Written (EN& FR) Speech (EN & FR)	OP portal static website	• Interruptions are functional • Reviewed tone • Improve static website design • Speech includes clearer guidance to the user
Sprint 6 (18/12)	All users stories	Written (EN& FR) Speech (EN & FR)	OP portal static website	• Implemented error messages • French tone has been improved

During each sprint the 9 acceptance criteria was used to evaluate the progress and update the backlog. Table 9 indicates the observed achievements and improvement points noticed across sprints.

TABLE 10: OVERALL SPRINT FINDINGS	
Acceptance criteria	**Description**
AC.1 Search Guidance	**Observed achievements** • Metadata used for filtering (author, date, type, subject & article language) • Can search within search after initial results • Good guidance, the bot is able to recognize the document search Intent for the user stories tested **Improvement points** • All entities are not recognised perfectly
AC.2 User Experience	**Observed achievements** • Assistant always answers • User can exit at any point • Clear messages and differentiated OPIA tone • Human-like triggers included (how are you? Etc.) • Interruptions in dialogue flow is functioning • Font size appropriate

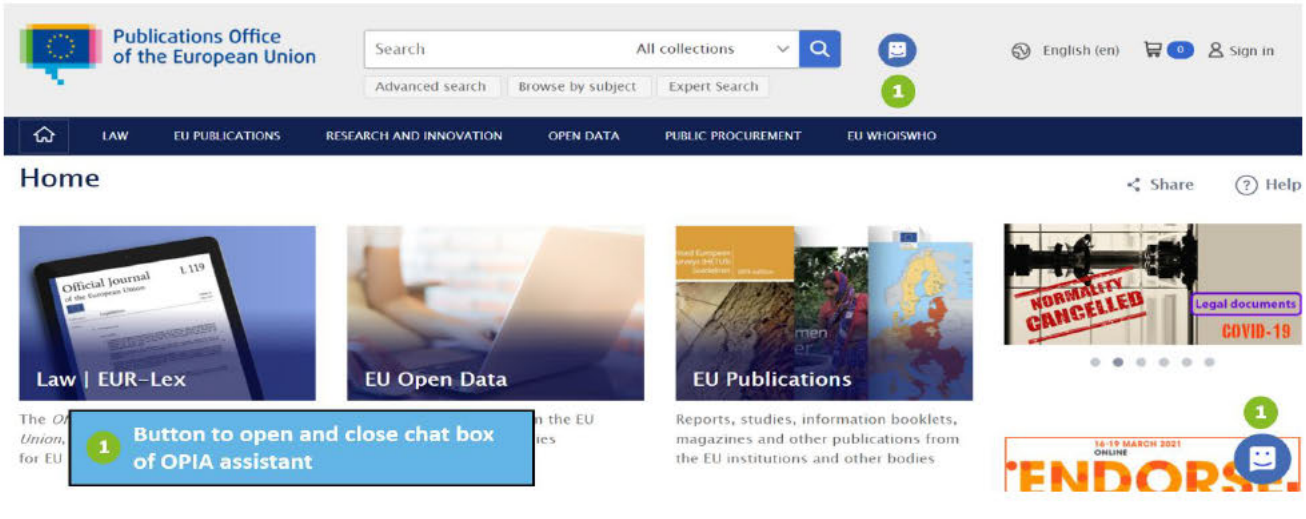
	Improvement points • Sensitivity of adaptive cards for tick boxes (to be improved in Ms version 1.3)
AC.3 Response speed	**Observed achievements** • All answers are readable in displayed box (size appropriate to facilitate reading) • Voice assist reading of messages n speech bot (typically longer messages) **Improvement points** • • Reminding user of available assistance after inactivity not built-in
AC.4 Conversational robustness	**Observed achievements** • Entities and Intents are mostly well recognized • All buttons are functioning • Analytics files have been used to improve the entity recognition • No dead ends in bot dialogue flow • OPIA clarifies when user is misunderstood • Voice bot adjusts words heard to closest match if not 100% correct **Improvement points** • Not all spelling mistakes are picked up as correct intents
AC.5 Interaction	**Observed achievements** • Adaptive cards remain active, thus user can reselect when they want to change selection • Function to scroll up and down to navigate through the conversation • Three dynamic random ordered search dialog flow **Improvement points** • Not assessing when user interacts with the button as the testing took place on static webpage • Automatic detection of user language and method to have one implemented assistant
AC.6 Speech-to-text functionality	**Observed achievements** • Speech-to-text able to understand voice input (EN & FR) • Bot able to read the response to the user aloud • Simpler speck bot with buttons implemented for speech. **Improvement points** • User with accent could be misunderstood • Voice command not able to convert an action to open final document
AC.7 Search efficiency	**Observed achievements** • OPIA only asks three questions before offering first view of the documents • Two options are proposed with option to see a further two options • Document title, icon and description provided where available • User can filter further after initial document proposal (search in search) **Improvement points** • Dynamic flow to be further improve by using API data

AC.8 Search conversion	**Observed achievements** • OPIA provides direct link to access articles • Assistant can refer user to advance search or the help page if further refinement is required **Improvement points** • No points listed during sprint evaluation
AC.9 Relevance of documents	**Observed achievements** • Documents proposed seem relevant to the search **Improvement points** • UAT foreseen for implementation

The assistant is implemented on a simulation of the OP Portal, using Azure web chat customization, to have a better idea of the final solution and to display certain functionalities such as clicking on the button to open the assistant. This service allows rich customization options to personalize the size, colours, buttons and font of the virtual assistant as designed in Stream B: Interface design.

Two buttons will be added on beta of the Publication Office website, one at the footer of the webpage always visible at an absolute position, and another one in the header of the webpage, at the right of the portal search box (see deliverable B.1 Prototype). These buttons, when clicked, will open the Publication Office Intelligent Assistant box at an absolute position at the bottom right of the webpage. The virtual assistant would be available through this window. Users can use it through the bottom message box at the bottom or by clicking on the microphone icon at the bottom right. The virtual assistant can be closed by clicking on two buttons, one at the bottom right at an absolute position, and another in the virtual assistant header at the right.

FIGURE 7: HOW THE VIRTUAL ASSISTANT WILL BE DISPLAY ON PUBLICATION OFFICE PORTAL



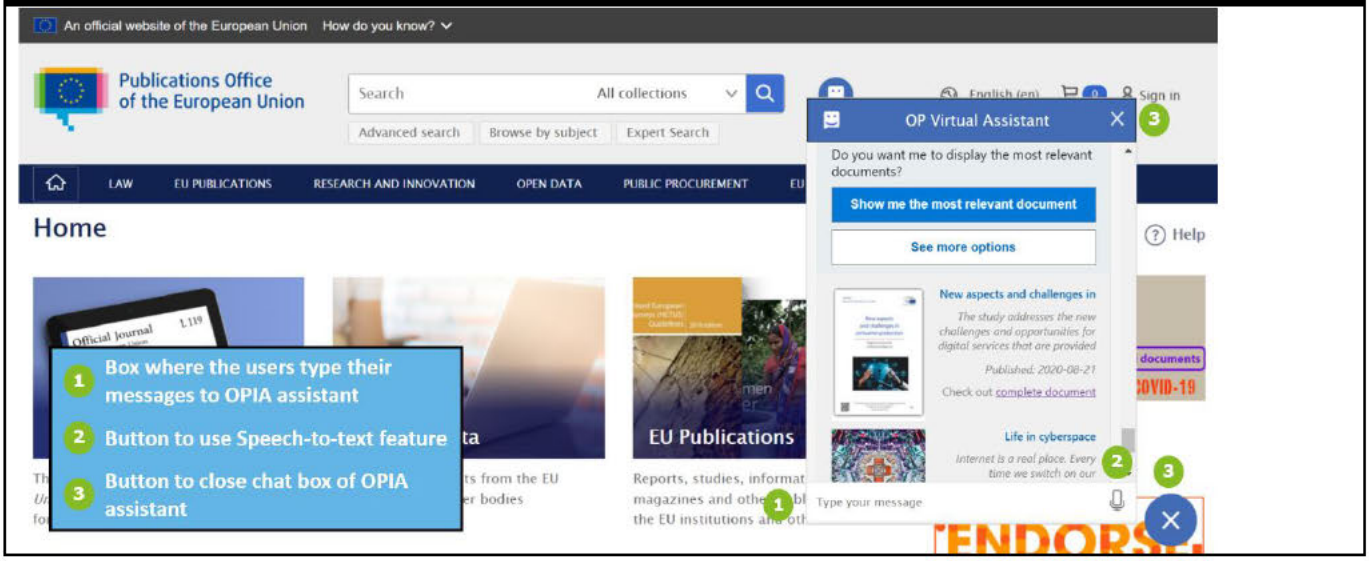
This visual version is for POC purposes only and full adaptations of the web pages will be done after POC phase. For the POC the OPIA should be displayed only on the home page and the search page, with the functionalities to open and close the OPIA by clicking on the icon and selecting the microphone to interact with voice capabilities.

Currently the bot is hosted on the Deloitte Azure tenant in a sandbox environment without the ambition to move towards production as full industrialization will take place in Publication Office Azure tenant with security measures to be defined as part of the industrialization approach (see deliverable C.2 Approach for industrialization). For the OPIA POC certain security measures have been implemented the Deloitte azure environment have IP filtering excluding external IP addresses from accessing and the environment is completely compliant with Deloitte Cyber Sec rules and regulations. Furthermore access to the static webpage is also monitored to the security measures such as IP filtering. Internal Microsoft Azure security features such as token generation and Microsoft Keys Vault also ensures the security of the POC development end to end. The main security goal is avoiding to share or to display any direct line keys or speech direct line keys to the front end. To do so, tokens represents an intermediate and temporary way to give authentication to users to access the virtual assistant. To add a new security layout, it is possible to gather and to store secret keys with Keys Vault on Azure to secure them and regenerate them automatically after a certain period.

4. Conclusion

This documentation described the technical functionalities of OPIA virtual assistant with the architecture,

FIGURE 8: HOW THE VIRTUAL ASSISTANT WILL BE DISPLAY ON PUBLICATION OFFICE PORTAL



conversation flow and the API calls. It details how the intelligent assistant recognises Intents and Entities with machine learning techniques from the user's input. The assistant was evaluated based on 9 acceptance criteria describing the conversational robustness, the guidance of the user, and other standards.

The following attention points should be addressed:

- **Automatic recognition of language**: At the moment the speech and written assistant is separated in two bots (both available in English and French, thus four bots in total). This proves the functioning of both features for POC purposes. For deployment purposes, one bot would need to automatically detect and update the flow based on the user's preference. Such detection is possible through middleware, which is a solution that will be investigated for the deployment of

the OPIA. A middleware is a software that acts as a bridge between a software and a network. In this particular case, we could foresee the bot starting with the English written version as a basis. The middleware will be used to detect if the user would like to use the speech version instead. For instance, once the user selects the microphone button the middleware would redirect the user to the speech version of the bot. Similarly, this would be applied to distinguish between the English and French Intelligent Assistant by detecting the language written by the user.

- **OPIA conversational flow:** Certain improvements flagged in the evaluation of the POC will be incorporated in the backlog stream for OPIA implementation. The list of challenges and the proposal to overcome these based on additional Azure components or further training of the LUIS model will be described in detail in the deliverable: proposal of industrialisation.

The next steps include the following:

- **Validation of OPIA POC:** As part of the delivery of the intelligent assistant, the OPIA POC will be displayed to the Publication office during a workshop. The three user stories will be illustrated and all acceptance criteria will be considered.
- **Showcase the OPIA at webinar:** The Publication Office plans to showcase the development of a search assistant at a public event. For this Deloitte will provide the presentation support to ensure the assistant functionalities are displayed and presented according to the two key functionalities and languages in scope.

5. Appendix

5.1 LUIS prebuilt packages

For the OPIA POC, 10 different intents were defined to guide the user. This stretches from a simple 'Greeting intent' that recognizes a greeting message to the most complex intent for searching documents that facilitates a structured search. This flow was selected for the POC to ensure that the goal of the search intent (to find a document) is reached by the user. Additional intents can be defined.

Each intents have been created from scratch for the POC scope but LUIS proposes several prebuilt intents and entities to save time and increase performance. These prebuilt intents and entities are only available for most popular language as the table 10 below shows:

TABLE 11: PREBUILT DOMAIN, INTENTS AND ENTITIES ON LUIS

Prebuilt packaging	Description	Include
Domain	Pre-trained models of intents and entities that work together for domains or common categories of client applications.	Intents, Utterances, entities
Intents	LUIS provides prebuilt intents and their utterances for each of its prebuilt domains. Intents can be added without adding the whole domain. Adding an intent is the process of adding an intent and its utterances to your app. Both the intent name and the utterance list can be modified.	Intents and utterances
Entities	LUIS includes a set of prebuilt entities for recognizing common types of information, like dates, times, numbers, measurements, and currency. Prebuilt entity support varies by the culture of your LUIS app. For a full list of the prebuilt entities that LUIS supports, including support by culture, see the prebuilt entity reference.	Entities only

From Microsoft Documentation, there are several prebuilt domains available in several language. Here's a non-exhaustive table of domains covered by LUIS:

TABLE 12: PREBUILT DOMAINS AVAILABLE ON LUIS

Prebuilt domain	Description
Calendar	The Calendar domain provides intent and entities for adding, deleting, or editing an appointment, checking participants availability, and finding information about a calendar event
Communication	Sending messages and making phone calls.
Email	Manage email tasks such as reading, replying and flagging emails
Note	Manage note taking tasks such as creating, editing, and finding notes.
Places	Handling queries related to places like businesses, institutions, restaurants, public spaces, and addresses.
To-do	Handling requests for task list.
Utilities	Handling common requests across domains, like 'help', 'repeat', 'start over.'
Weather	Getting weather reports and forecasts.
Web	Navigating to a website.
Restaurant Reservation	Manage tasks associated with reservations such as book a table at a restaurant, changing or cancelling a reservation at a restaurant
Home Automation	Controlling smart home devices like lights and appliances

5.2 List of all Intents and trained Utterances

TABLE 13: INTENT DEFINITIONS	
English	**French**
Foreign	
• ciao • cześć • dzień dobry • god dag • goddag • goedendag • guten tag • hallo • halløj • hej • hoi • hola • konnichiwa • merhaba • oi • olá • privet • salve • selam • shalom • tjena • witaj • yassas • yassou • zdravstvuyte • asalaam alaikum	• asalaam alaikum • ciao • cześć • dzień dobry • god dag • goddag • goedendag • guten tag • hallo • halløj • hej • hoi • hola • konnichiwa • merhaba • oi • olá • privet • salve • selam • shalom • tjena • witaj • yassas • yassou • zdravstvuyte • ¿qué tal?
Greetings	
• good afternoon • good morning • good night • halo • hello • hey • hey there • hi op bot • hi there • it's a pleasure meeting you • it's an honour • morning • my pleasure • nice to meet you	• bienvenue • bjr • bonjour • bonjour à toi • bonne journée • bonsoir • commence • coucou • début • enchanté • hello • hey • hi • quel honneur de te rencontrer

• night • now • start • today • what's up • what’s up • yo • afternoon • begin • evening	• quel honneur de vous rencontrer • salut • salut toi • slt • yo
Help	
• can I ask you for a favour • can I get assistance • can I get help • can you assist me • can you give me a hand • can you help me • can you help me with something • can you please assist me • can you please do something for me • can you please help • can you provide some help • could I bother you • could you assist me • could you help me out • do you know how to • do you think you can help • help • help me please • how am I suppose to search a document with op assistant • I am asking for help • I need help • I need you • I wonder if you can help • is it possible to assist me • lend me a hand please • please help • what are you supposed to do • what do you do • what is this bot for • what is your purpose • would you mind assisting me • would you mind to help • would you offer some help • can I ask you a favour	• ca te derangerait de m'aider un peu • j'ai besoin d'aide • j'ai besoin de ton aide • j'ai besoin de ton soutien • je me demande si tu peux m'aider • je peux te déranger stp • peux tu etre dérangé • peux tu m'accorder de ton temps • peux tu m'assister • peux tu me rendre un service • peux-tu m'aider • peux-tu m'apporter de l'aide • pourrais je avoir ton aide • s'il te plait aide moi • soutiens moi • stp aide moi • svp aidez moi • tu dois me venir en aide • tu penses pouvoir m'aider • tu peux m'accorder un peu de ton temps • tu peux m'aider • tu peux me venir au secours • tu peux me venir en aide • tu peux m'offrir ton aide • tu sais comment m'aider • viens m'aider • apporte moi de l’aide
Mood	

• are you ok • are you well • do you feel good • do you feel great • how are you • how are you doing • how are you feeling • how do you do • how do you feel • how is it going • how is life • how's your day • what is going on • what is your mood • what's going on • what's new • you alright • you ok • are you fine	• ca va bien • comment ca se passe • comment ca va • comment se passe ta journée • comment tu te sens • comment va la vie • comment vas tu • quoi de neuf • quoi de nouveau • tout est bon • tout est ok • tout marche comme sur des roulettes • tout roule • tout se passe bien • tout va bien • tu te sens bien • tu vas bien • ca va
Name	
• can you repeat your name • do you have a name • how can I call you • how do people call you • how do we call you • how do we name you • how may I adress you • how should I call you • may I ask your name • may I have your name • present yourself • say your name • tell me your name • what are you named • what do you call yourself • what is your name • what's your name • who are you • would you mind saying your name • are doing well	• c'est quoi ton prénom • comment devrais-je t'appeler • comment tu t'appelles • comment tu te prenommes • comment veux-tu que je te prénomme • dis moi ton nom • dis moi ton prénom • je peux t'appeler comment • on te surnomme comment • presente toi • t'es qui • tu as un nom • tu as un prénom • tu as un surnom • tu es qui • tu peux me rappeler ton nom • tu t'appelles comment • c'est quoi ton nom
Search	
• {@ArticleTheme=agricultural} • {@ArticleTheme=agriculture} • {@ArticleTheme=agronomy} • {@ArticleTheme=ai} • any info related to {@ArticleTheme=data collection}	• {@ArticleTheme=ai} • {@ArticleTheme=artificial intelligence} • {@ArticleTheme=biology} • {@ArticleTheme=biosphère} • {@ArticleTheme=blockchain}

- {@ArticleTheme=biodiversity}
- {@ArticleTheme=biology}
- blockchain
- can I have en document
- can I have English document
- can I have fr document
- can I have French document
- can I have some help to find a report on {@ArticleTheme=data privacy regulations}
- can I have some help to find a report on {@ArticleTheme=women's rights}
- can I receive a document discussing the topic of {@ArticleTheme=single market in Europe}
- can I receive an article about {@ArticleTheme=environmental taxes}
- can I receive an article talking about {@ArticleTheme=environmental policies}
- can you bring some help by searching for a document about {@ArticleTheme=energy efficiency}
- can you find a doc about {@ArticleTheme=data privacy}
- can you find a doc about {@ArticleTheme=immigration}
- can you find a doc about {@ArticleTheme=sustainability}
- can you give me a document providing any information on {@ArticleTheme=economic social cultural rights}
- can you help me finding any information about {@ArticleTheme=consumer protection in e commerce}
- can you help me finding some reports talking about {@ArticleTheme=information security}
- can you help me looking for an article discussing about {@ArticleTheme=brexit}
- can you help me searching for an article about {@ArticleTheme=sustainability transition}
- can you look for a document about {@ArticleTheme=foreign policy}
- can you look for an article talking about {@ArticleTheme=open data portal}
- can you look for any file in the search portal about {@ArticleTheme=data collection}
- {@ArticleTheme=bouleversement météorologique}
- {@ArticleTheme=changement climatique}
- {@ArticleTheme=changement de climat}
- cherche-moi un document qui met en avant {@ArticleTheme=l'inégalités des sexes}
- cherche-moi un document qui met en avant {@ArticleTheme=protection de l'environnement}
- cherche-moi un dossier en relation avec {@ArticleTheme=protection de l'environnement}
- cherche-moi un dossier qui discute de {@ArticleTheme=règlementations environnementales}
- cherche-moi un journal au sujet de {@ArticleTheme=collection de données}
- cherche-moi un journal qui discute de des {@ArticleTheme=disparités dans la société}
- cherche-moi un rapport discutant de {@ArticleTheme=information confidentielle}
- cherche-moi une information qui discute de des {@ArticleTheme=disparités dans la société}
- cherche-moi une information qui discute de {@ArticleTheme=données confidentiels}
- cherche-moi une information qui discute de {@ArticleTheme=fonte des glaces}
- cherche-moi une information qui discute de {@ArticleTheme=l'égalité des genres}
- cherche-moi une revue mettant en avant {@ArticleTheme=collection de données}
- cherche-moi une revue qui discute de le {@ArticleTheme=droit d'expression}
- {@ArticleTheme=climat}
- {@ArticleTheme=collection de données}
- {@ArticleTheme=combat contre le réchauffement climatique}
- {@ArticleTheme=déforestation}
- Des {@ArticleTheme=disparités dans la société}
- {@ArticleTheme=données}
- {@ArticleTheme=données confidentielles}
- {@ArticleTheme=données privées}

- can you provide me a document with information about {@ArticleTheme=equality of rights in eu}
- can you provide me a document with information about {@ArticleTheme=the impact of ai on data privacy}
- can you provide me information about {@ArticleTheme=environmental tax}
- can you provide me information about {@ArticleTheme=impact of blockchain}
- can you provide me information about {@ArticleTheme=woman's right}
- can you search for articles related to {@ArticleTheme=open data}
- can you search for docs having info on {@ArticleTheme=European wealth data}
- can you send me a doc about {@ArticleTheme=freedom of movement}
- can you send me a doc about {@ArticleTheme=personal data protection}
- can you send me a doc in en about {@ArticleTheme=environmental regulations in eu}
- can you try to find an article with information on {@ArticleTheme=offshore installations}
- {@ArticleTheme=citizen rights in eu}
- {@ArticleTheme=citizenship}
- {@ArticleTheme=civil liberties}
- {@ArticleTheme=civil rights}
- {@ArticleTheme=climate change}
- {@ArticleTheme=climate change policies}
- {@ArticleTheme=climate change policy}
- {@ArticleTheme=climate environment}
- {@ArticleTheme=climate smart agriculture}
- {@ArticleTheme=cloud computing}
- {@ArticleTheme=computer security}
- {@ArticleTheme=constitution}
- {@ArticleTheme=constitutional rights}
- {@ArticleTheme=consumer law}
- {@ArticleTheme=consumer protection}
- {@ArticleTheme=consumer protection in e commerce}
- {@ArticleTheme=consumer rights}
- {@ArticleTheme=consumer rights and protection}
- {@ArticleTheme=écologie}
- {@ArticleTheme=écologie}
- {@ArticleTheme=effet de serre}
- {@ArticleTheme=Energie renouvelables}
- {@ArticleTheme=Energie verte}
- {@ArticleTheme=environnement}
- essaie de trouver un article qui discute de des {@ArticleTheme=disparités dans la société}
- Essaie de trouver un article qui discute d'énergie renouvelable
- Essaie de trouver un document en relation avec {@ArticleTheme=impact des déchets}
- Essaie de trouver un document en relation avec la {@ArticleTheme=loi sociale}
- Essaie de trouver un document en relation avec {@ArticleTheme=l'inégalités des sexes}
- Essaie de trouver un dossier en relation avec {@ArticleTheme=l'inégalité sociale}
- Essaie de trouver un journal parlant de {@ArticleTheme=données confidentielles}
- Essaie de trouver une information à propos de {@ArticleTheme=données privés}
- Essaie de trouver une information discutant de {@ArticleTheme=énergie renouvelable}
- Essaie de trouver une information en relation avec {@ArticleTheme=protection de l'environnement}
- Essaie de trouver une revue qui parle de {@ArticleTheme=données}
- Essaye de chercher un article discutant de des {@ArticleTheme=disparités dans la société}
- Essaye de chercher un article discutant de {@ArticleTheme=gaspillages}
- Essaye de chercher un article discutant de le {@ArticleTheme=droit du consommateur}
- Essaye de chercher un document parlant de {@ArticleTheme=biologie}
- Essaye de chercher une revue mettant en avant {@ArticleTheme=portail de données}
- {@ArticleTheme=fonte des glaces}
- {@ArticleTheme=gaspillages}
- {@ArticleTheme=gdpr}

- corporate governance
- {@ArticleTheme=cyber}
- {@ArticleTheme=cybercrime}
- {@ArticleTheme=cyber law}
- {@ArticleTheme=cybersecurity}
- {@ArticleTheme=cyberterrorism}
- {@ArticleTheme=data}
- {@ArticleTheme=data collection}
- {@ArticleTheme=data management}
- {@ArticleTheme=data privacy}
- {@ArticleTheme=data privacy regulations}
- {@ArticleTheme=data processing}
- {@ArticleTheme=data processing-regulation}
- {@ArticleTheme=data protection}
- {@ArticleTheme=data protections}
- {@ArticleTheme=data rights}
- {@ArticleTheme=data security}
- {@ArticleTheme=data store}
- {@ArticleTheme=data technology}
- {@ArticleTheme=deforestation}
- {@ArticleTheme=democracy}
- {@ArticleTheme=ecological}
- {@ArticleTheme=ecology}
- {@ArticleTheme=economic social cultural rights}
- {@ArticleTheme=ecosystems}
- {@ArticleTheme=energy}
- {@ArticleTheme=environment}
- {@ArticleTheme=environmental}
- {@ArticleTheme=environmental policies}
- {@ArticleTheme=environmental taxes}
- {@ArticleTheme=equality of rights in eu}
- Eu {@ArticleTheme=environnemental policy}
- {@ArticleTheme=European climate change programme}
- {@ArticleTheme=European commission}
- {@ArticleTheme=European union environmental regulation}
- {@ArticleTheme=farming}
- {@ArticleTheme=foreign policy}
- {@ArticleTheme=free speech}
- {@ArticleTheme=freedom of citizens}
- {@ArticleTheme=freedom of religion}
- {@ArticleTheme=freedom of speech}
- {@ArticleTheme=freedoms}
- {@ArticleTheme=gouvernance des données}
- Il faut que je déniche un article en relation avec {@ArticleTheme=gdpr}
- Il faut que je déniche un dossier qui met en avant {@ArticleTheme=bouleversement météorologique}
- Il faut que je déniche un dossier qui met en avant là {@ArticleTheme=citoyenneté}
- Il faut que je déniche un dossier qui met en avant {@ArticleTheme=l'inégalités des sexes}
- Il faut que je déniche un journal parlant de {@ArticleTheme=données sécurisés}
- Il faut que je déniche un journal qui met en avant {@ArticleTheme=l'impact du brexit sur l'UE}
- Il faut que je déniche un rapport concernant {@ArticleTheme=données sécurisées}
- Il faut que je déniche un rapport qui met en avant {@ArticleTheme=l'inégalités des sexes}
- Il faut que je déniche une information qui parle de {@ArticleTheme=l'égalité des genres}
- Il faut que je déniche une information qui parle de {@ArticleTheme=règlementations environnementales}
- Il faut que je déniche une revue qui discute de {@ArticleTheme=règlementations environnementales}
- Il faut que je déniche une revue qui met en avant {@ArticleTheme=écologie}
- Il faut que je déniche une revue qui met en avant {@ArticleTheme=information confidentielle}
- Il me faut un document qui discute de {@ArticleTheme=rgpd}
- Il me faut un dossier parlant de la {@ArticleTheme=citoyenneté}
- Il me faut un dossier parlant de l' {@ArticleTheme=inégalité sociale}
- Il me faut un dossier qui met en avant {@ArticleTheme=écologie}
- Il me faut un journal qui met en avant {@ArticleTheme=écologie}

- {@ArticleTheme=gdpr}
- {@ArticleTheme=global warming}
- {@ArticleTheme=green economy}
- {@ArticleTheme=green energy}
- {@ArticleTheme=greenhouse gas}
- {@ArticleTheme=health care}
- {@ArticleTheme=health inequalities in Europe}
- hello, may I ask you to find me {@ArticleTheme=green energy} file in fr
- help me finding a document with information about new tech in the {@ArticleTheme=energy market}
- help me to find an article about the {@ArticleTheme=electricity market}
- {@ArticleTheme=human rights}
- I am interested in {@ArticleTheme=clean energy}
- I am interested in knowing about {@ArticleTheme=privacy rights} in the eu
- I am interested in knowing more about the {@ArticleTheme=consequence of the brexit}
- I am looking for documents related to {@ArticleTheme=data security}
- I am searching for a document
- I am searching for a document in English about {@ArticleTheme=protection of environment}
- I am searching for a document in French about {@ArticleTheme=consumer law}
- I am searching for {@ArticleTheme=artificial intelligence} docs
- I am searching for {@ArticleTheme=data protections} related documents
- I am searching for French {@ArticleTheme=consumer rights} pdf
- I am trying to find a report about {@ArticleTheme=eu environmental climate regulations}
- I am trying to find an information about {@ArticleTheme=citizen rights}
- I am trying to find an information in French concerning {@ArticleTheme=migrants rights in eu}
- I am trying to find information concerning {@ArticleTheme=data privacy}
- Il me faut un journal qui parle de {@ArticleTheme=énergie verte}
- Il me faut un rapport discutant de {@ArticleTheme=données privées}
- Il me faut un rapport qui parle de les {@ArticleTheme=lois étrangères}
- Il me faut une revue à propos de {@ArticleTheme=données privées}
- Il me faut une revue qui parle de {@ArticleTheme=effet de serre}
- Il me faut une revue qui parle de {@ArticleTheme=l'égalité des genres}
- Il me faut une revue qui parle de les {@ArticleTheme=migrants}
- {@ArticleTheme=impact des déchets}
- {@ArticleTheme=information confidentielle}
- {@ArticleTheme=intelligence artificielle}
- J'ai besoin que tu me trouves un article discutant de {@ArticleTheme=ai}
- J'ai besoin que tu me trouves un document qui parle de {@ArticleTheme=déforestation}
- J'ai besoin que tu me trouves un dossier au sujet de {@ArticleTheme=collection de données}
- J'ai besoin que tu me trouves un rapport concernant {@ArticleTheme=déforestation}
- J'ai besoin que tu me trouves un rapport qui discute de le {@ArticleTheme=droit d'expression}
- J'ai besoin que tu me trouves un rapport qui discute de les {@ArticleTheme=lois étrangères}
- J'ai besoin que tu me trouves un rapport qui met en avant {@ArticleTheme=ai}
- J'ai besoin que tu me trouves une information à propos de le {@ArticleTheme=droit du citoyen}
- Je dois chercher un article au sujet de la {@ArticleTheme=précarité}
- Je dois chercher un document qui met en avant {@ArticleTheme=ressource de données}
- Je dois chercher un journal mettant en avant {@ArticleTheme=l'intelligence artificielle}

• I have to find a document in relation to {@ArticleTheme=climate change policies} • I have to find a document in relation to {@ArticleTheme=European union environmental regulations} • I have to find a file in relation to {@ArticleTheme=climate change policies} • I have to find a file in relation to {@ArticleTheme=human rights} • I have to find an article related to {@ArticleTheme=data collection} • I have to search for an article in English about {@ArticleTheme=general data protection} • I look for an article about the environment on the search portal • I look for an article on the search portal about {@ArticleTheme=consumer rights} • I look for an article on the search portal about {@ArticleTheme=data governance} • I look for articles about {@ArticleTheme=environmental policies} • I look for articles about {@ArticleTheme=European commission} • I look for articles about {@ArticleTheme=ml} • I need a doc in fr related to {@ArticleTheme=data usage} • I need a document in English about {@ArticleTheme=blockchain} • I need a fr article in relation to {@ArticleTheme=blockchain} • I need a French doc related to {@ArticleTheme=single market} • I need a report in French about {@ArticleTheme=climate change programme} • I need a report in French about {@ArticleTheme=data protection} in the eu • I need a report in French about {@ArticleTheme=foreign policy} • I need an English report dealing with {@ArticleTheme=blockchain} • I need an English report dealing with {@ArticleTheme=foreign policy}	• Je dois chercher un journal qui met en avant {@ArticleTheme=changement climatique} • Je dois chercher un rapport qui met en avant {@ArticleTheme=pollution} • Je dois chercher un rapport qui parle de {@ArticleTheme=énergie verte} • Je dois chercher une revue discutant de le {@ArticleTheme=droit d'expression} • Je dois chercher une revue discutant de les {@ArticleTheme=migrants} • Je dois chercher une revue en relation avec {@ArticleTheme=intelligence artificielle} • Je dois chercher une revue qui parle de {@ArticleTheme=énergie verte} • Je dois trouver un article qui met en avant {@ArticleTheme=pollution} • Je dois trouver un document concernant le {@ArticleTheme=droit des femmes} • Je dois trouver un journal au sujet de {@ArticleTheme=pollution} • Je dois trouver un journal en relation avec {@ArticleTheme=l'impact du brexit sur l'UE} • Je dois trouver un journal en relation avec {@ArticleTheme=l'inégalité sociale} • Je dois trouver un journal qui discute de {@ArticleTheme=artificial intelligence} • Je dois trouver un rapport parlant de {@ArticleTheme=artificial intelligence} • Je dois trouver une information qui parle de {@ArticleTheme=données} • Je veux que tu trouves un article à propos de {@ArticleTheme=taxe environnementale} • Je veux que tu trouves un article qui met en avant {@ArticleTheme=la précarité} • Je veux que tu trouves un article qui met en avant {@ArticleTheme=l'impact du brexit sur l'UE} • Je veux que tu trouves un article qui parle de {@ArticleTheme=ml} • Je veux que tu trouves un document en relation avec {@ArticleTheme=ml} • Je veux que tu trouves un dossier mettant en avant le {@ArticleTheme=droit de l'homme}

- I need help looking for a report in English about {@ArticleTheme=nature conservation}
- I want a doc on {@ArticleTheme=social inequality}
- I want a file about {@ArticleTheme=information security}
- I want a report about {@ArticleTheme=personal information security}
- I want a report in fr about {@ArticleTheme=hr}
- I want {@ArticleTheme=ai} docs
- I want an English doc about {@ArticleTheme=citizen rights}
- I want articles related to {@ArticleTheme=consumer rights and protection}
- I want {@ArticleTheme=data technology} docs in French
- I want docs specialized in {@ArticleTheme=social inequalities}
- I want information in relation to {@ArticleTheme=gdpr}
- I want to receive info related to {@ArticleTheme=sustainability impact}
- I would like a doc about {@ArticleTheme=eu commission}
- I would like a doc related to {@ArticleTheme=environmental changes in eu}
- I would like a report related to {@ArticleTheme=social inequality}
- I would like any article in English discussing about {@ArticleTheme=climate smart agriculture}
- I would like to receive some help to find a doc related to {@ArticleTheme=offshore} topic
- I would like to receive some help to find a doc related to the {@ArticleTheme=gdpr}
- I would need your help to search for a file about {@ArticleTheme=climate environment}
- I'm interested in finding information on {@ArticleTheme=data security} in eu
- I'm looking for a paper discussing about {@ArticleTheme=climate change policy}

- Je veux que tu trouves une information au sujet de {@ArticleTheme=taxe environnementale}
- Je veux que tu trouves une revue concernant {@ArticleTheme=données sécurisés}
- Je veux un article concernant changement de {@ArticleTheme=climat}
- Je veux un article concernant {@ArticleTheme=open data portal}
- Je veux un article qui parle de {@ArticleTheme=données}
- Je veux un article qui parle de {@ArticleTheme=l'égalité des genres}
- Je veux un document au sujet de {@ArticleTheme=machine learning}
- Je veux un document qui discute de {@ArticleTheme=machine learning}
- Je veux un document qui parle de le {@ArticleTheme=droit des femmes}
- Je veux un document qui parle de les {@ArticleTheme=lois étrangères}
- Je veux un journal à propos de {@ArticleTheme=données privées}
- Je veux une information parlant de {@ArticleTheme=climat}
- Je veux une revue parlant de l'{@ArticleTheme=immigration}
- Je voudrais que tu cherches un article qui parle de {@ArticleTheme=déforestation}
- Je voudrais que tu cherches un document à propos de {@ArticleTheme=réglementations écologiques}
- Je voudrais que tu cherches un document au sujet de {@ArticleTheme=collection de données}
- Je voudrais que tu cherches un document concernant {@ArticleTheme=l'utilisation des données}
- Je voudrais que tu cherches un document parlant de la {@ArticleTheme=citoyenneté}
- Je voudrais que tu cherches un journal discutant de le {@ArticleTheme=droit du consommateur}

- I’m looking for an article talking about {@ArticleTheme=ai} and {@ArticleTheme=ml}
- I’m searching for a doc related to {@ArticleTheme=data usage rights}
- I’m trying to collect information about {@ArticleTheme=European union environmental regulations}
- I’m trying to find a file showing the {@ArticleTheme=consequence of climate change}
- I’m trying to find an information in relation to {@ArticleTheme=the right to reuse data}
- I’m trying to look for a document showing {@ArticleTheme=sustainability impact assessment}
- I’m trying to look for an article showing {@ArticleTheme=retail regulations}
- I’m trying to search for docs about {@ArticleTheme=equality of rights}
- I’m trying to search for docs concerning {@ArticleTheme=data privacy} topic
- {@ArticleTheme=immigration}
- {@ArticleTheme=impact of ai on education}
- {@ArticleTheme=information security}
- {@ArticleTheme=information technology}
- {@ArticleTheme=intellectual property}
- is there a doc about {@ArticleTheme=big data}
- is there a doc about gdpr in the search portal
- is there a document talking about {@ArticleTheme=data processing-regulation} on the search portal
- is there a file providing information on the {@ArticleTheme=European comission}
- is there an article about {@ArticleTheme=sustainability}
- is there any article talking about {@ArticleTheme=migration} on the search portal
- is there any doc about {@ArticleTheme=European data}
- is there any document having as a main topic {@ArticleTheme=green economy}
- Je voudrais que tu cherches un rapport mettant en avant {@ArticleTheme=intelligence artificielle}
- Je voudrais que tu cherches un rapport qui parle de {@ArticleTheme=environnement}
- Je voudrais que tu cherches une information au sujet de la {@ArticleTheme=précarité}
- Je voudrais que tu cherches une information au sujet de le {@ArticleTheme=droit du citoyen}
- Je voudrais que tu cherches une information qui met en avant {@ArticleTheme=utilisation des données}
- Je voudrais trouver un article qui met en avant {@ArticleTheme=changement climatique}
- Je voudrais trouver un journal qui parle de {@ArticleTheme=blockchain}
- Je voudrais un PDF concernant les {@ArticleTheme=énergies renouvelables} écrit en 2020 par la commission européenne
- La {@ArticleTheme=citoyenneté}
- {@ArticleTheme=la gouvernance des données}
- La {@ArticleTheme=loi sociale}
- La {@ArticleTheme=précarité}
- {@ArticleTheme=la sécurité de données}
- Le {@ArticleTheme=droit de l'homme}
- Le {@ArticleTheme=droit des femmes}
- Le {@ArticleTheme=droit d'expression}
- Le {@ArticleTheme=droit du citoyen}
- Le {@ArticleTheme=droit du consommateur}
- {@ArticleTheme=l'égalité des genres}
- Les {@ArticleTheme=lois étrangères}
- Les {@ArticleTheme=migrants}
- L'{@ArticleTheme=immigration}
- {@ArticleTheme=l'impact du brexit sur l’UE}
- {@ArticleTheme=l'inégalité des sexes}
- {@ArticleTheme=l'inégalité sociale}
- {@ArticleTheme=l'intelligence artificielle}
- {@ArticleTheme=l'utilisation des données}
- Ma recherche concerne un rapport qui discute de {@ArticleTheme=effet de serre}

- is there any file giving information on {@ArticleTheme=eu citizenship}
- is there any report talking about {@ArticleTheme=gender inequality}
- is there any report talking about {@ArticleTheme=open data}
- is there any report talking about {@ArticleTheme=renewable energy in eu}
- is there in the search portal a file talking about {@ArticleTheme=health inequalities}
- I've to find a report covering {@ArticleTheme=ml} topic
- look for a doc about {@ArticleTheme=impact of ai on education}
- look for a document related to {@ArticleTheme=citizen rights}
- look for a document related to {@ArticleTheme=data security}
- look for a document related to {@ArticleTheme=renewable sources}
- look for a file on the search portal about {@ArticleTheme=climate change}
- look for article concerning {@ArticleTheme=personal data} topic
- {@ArticleTheme=migration}
- my research concerns {@ArticleTheme=data technology} topic
- my research concerns info {@ArticleTheme=security} subject
- my research is about {@ArticleTheme=retail regulations}
- my research is based on finding a document about {@ArticleTheme=social law in eu}
- my research is related to {@ArticleTheme=eu environmental policy}
- {@ArticleTheme=nature conversation}
- {@ArticleTheme=network security}
- {@ArticleTheme=offshore installations}
- {@ArticleTheme=open data portal}
- {@ArticleTheme=overgrazing}
- {@ArticleTheme=personal information security}
- please help me finding a document about {@ArticleTheme=freedom of speech}
- Ma recherche concerne une revue qui met en avant {@ArticleTheme=utilisation des données}
- {@ArticleTheme=machine learning}
- {@ArticleTheme=ml}
- {@ArticleTheme=panneaux photovoltaïque}
- Peux-tu me trouver un article à propos de {@ArticleTheme=panneaux photovoltaïque}
- Peux-tu me trouver un document mettant en avant {@ArticleTheme=changement de climat}
- Peux-tu me trouver un journal concernant {@ArticleTheme=open data portal}
- Peux-tu me trouver un journal mettant en avant le {@ArticleTheme=droit de l'homme}
- Peux-tu me trouver un journal mettant en avant le {@ArticleTheme=droit du consommateur}
- Peux-tu me trouver une information à propos de {@ArticleTheme=planète verte}
- Peux-tu me trouver une information en relation avec {@ArticleTheme=ml}
- Peux-tu me trouver une revue discutant de {@ArticleTheme=fonte des glaces}
- Peux-tu chercher un document concernant {@ArticleTheme=réchauffement climatique}
- Peux-tu chercher un document discutant de {@ArticleTheme=ai}
- Peux-tu chercher un dossier parlant de {@ArticleTheme=impact des déchets}
- Peux-tu chercher une information concernant {@ArticleTheme=changement de climat}
- Peux-tu chercher une revue à propos de le {@ArticleTheme=droit du citoyen}
- Peux-tu chercher une revue à propos de l'{@ArticleTheme=immigration}
- Peux-tu chercher une revue au sujet de {@ArticleTheme=la gouvernance des données}
- {@ArticleTheme=planète verte}
- {@ArticleTheme=pollution}
- {@ArticleTheme=portail de données}

- please help me finding an article about {@ArticleTheme=sustainable development goals}
- please help me searching for a doc providing info about {@ArticleTheme=climate policies in eu}
- please help me searching for a report having information about {@ArticleTheme=European climate change programme}
- please look for an article discussing about {@ArticleTheme=health inequalities in Europe}
- please search a file concerning {@ArticleTheme=electricity market}
- please search a file concerning {@ArticleTheme=general regulations}
- please search a file concerning {@ArticleTheme=open data portal}
- please search for a file about {@ArticleTheme=freedom of expression}
- please search for a report about {@ArticleTheme=renewable energy}
- please search for an article in en about {@ArticleTheme=reusable data}
- pls help me searching for a report in English dealing with {@ArticleTheme=data collection}
- pls I need a doc about {@ArticleTheme=environmental taxes}
- pls I would like any report related to {@ArticleTheme=ai on education}
- pls look for an article about {@ArticleTheme=gdpr} in eu
- pls look for an article showing the impact of {@ArticleTheme=climate change}
- pls look for an article showing the {@ArticleTheme=impact of migration}
- pls search docs about {@ArticleTheme=social laws}
- {@ArticleTheme=polluting}
- {@ArticleTheme=pollution}
- {@ArticleTheme=privacy}
- {@ArticleTheme=privacy rights}
- {@ArticleTheme=property rights}
- {@ArticleTheme=protection of environment}
- {@ArticleTheme=renewable energy}
- {@ArticleTheme=protection de l'environnement}
- {@ArticleTheme=réchauffement climatique}
- {@ArticleTheme=réglementations écologiques}
- {@ArticleTheme=règlementations environnementales}
- {@ArticleTheme=ressources de données}
- {@ArticleTheme=rgpd}
- {@ArticleTheme=sécurité des données}
- Stp je voudrais un document qui parle de {@ArticleTheme=environnement}
- Stp je voudrais un rapport au sujet de {@ArticleTheme=gouvernance des données}
- {@ArticleTheme=taxe environnementale}
- Trouve moi un article au sujet de {@ArticleTheme=climat}
- Trouve moi un article parlant de {@ArticleTheme=artificial intelligence}
- Trouve moi un document au sujet de {@ArticleTheme=taxe environnementale}
- Trouve moi un document concernant {@ArticleTheme=données sécurisés}
- Trouve moi un dossier concernant le {@ArticleTheme=droit de l'homme}
- Trouve moi un dossier concernant le {@ArticleTheme=droit des femmes}
- Trouve moi un dossier qui parle de {@ArticleTheme=blockchain}
- Trouve moi un rapport en relation avec la {@ArticleTheme=loi sociale}
- Trouve moi une information à propos de {@ArticleTheme=blockchain}
- Trouve moi une information discutant de les {@ArticleTheme=migrants}
- Trouve moi une information en relation avec {@ArticleTheme=biologie}
- Trouve moi une information mettant en avant {@ArticleTheme=gaspillages}
- {@ArticleTheme=utilisation des données}
- Y a-t-il un article mettant en avant {@ArticleTheme=biosphère}
- Y a-t-il un dossier qui discute de {@ArticleTheme=machine learning}
- Y a-t-il un rapport parlant de l'{@ArticleTheme=immigration}

• {@ArticleTheme=retail regulations} • {@ArticleTheme=rights of citizenship} • search • search an article about {@ArticleTheme=right of citizen} • search for a doc with info about {@ArticleTheme=renewable technologies} • search for an article about {@ArticleTheme=machine learning} • search for {@ArticleTheme=ml} docs pls • searching for an article showing {@ArticleTheme=the importance of ml} • searching for article • searching for document • searching for document in {@ArticleTheme=ecology} • searching for {@ArticleTheme=ecology} document • searching for files talking about {@ArticleTheme=open data portal} • searching for {@ArticleTheme=green energy} document • {@ArticleTheme=security} • send me a doc discussing about {@ArticleTheme=renewable sources} • {@ArticleTheme=single market in Europe} • {@ArticleTheme=social inequality} • {@ArticleTheme=social law in eu} • {@ArticleTheme=sustainability transition} • {@ArticleTheme=sustainable development goals} • {@ArticleTheme=sustainable impact assessment} • the {@ArticleTheme=consequence of climate change} • {@ArticleTheme=the consequence of the brexit} • the {@ArticleTheme=impact of ai on data privacy} • {@ArticleTheme=the importance of ml} • {@ArticleTheme=the right to reuse data} {@ArticleTheme=voting rights}	• Y a-t-il un document mettant en avant {@ArticleTheme=combat contre le réchauffement climatique} • Y a-t-il un journal en relation avec {@ArticleTheme=bouleversement météorologique} • Y a-t-il un rapport à propos de {@ArticleTheme=la sécurité de données} y a-t-il un rapport parlant de la loi sociale

5.3 Entity types and list

All authors metadata can be detected with a pattern. Indeed, by defining where they stand in a sentence, every possible authors would be detected if they are well placed. Same for the date metadata.

TABLE 14: ENTITIES FOR THE OPIA PROJECT

Entities	Type	Examples in English	Examples in French
Article author	Pattern.any	• [in {ArticleYear} [written by {ArticleAuthor}]] • [written by {ArticleAuthor} [in {ArticleYear}]]	• [en {ArticleYear} [(de|par) {ArticleAuthor}]] • [écrit par {ArticleAuthor} [en {ArticleYear}]]
Article language	List	• French • English	• Francais • Anglais
Article theme	Machine learned	• See separate table with user stories	• See separate table with user stories
Article type	List	• Doc • Ebook • Html • Image • Multimedia • Pdf • Postcard • Poster • Print • Xml	• Doc • Ebook • Html • Image • Multimedia • Pdf • Postcard • Poster • Print • Xml
Article year	Pattern.any	• [in {ArticleYear} [written by {ArticleAuthor}]] • [written by {ArticleAuthor} [in {ArticleYear}]]	• [en {ArticleYear} [(de|par) {ArticleAuthor}]] • [écrit par {ArticleAuthor} [en {ArticleYear}]]

5.4 Entities by user story

TABLE 15: ALL ARTICLE THEME ENTITIES		
User story 1 - GDPR	**User story 2 – Consumer rights**	**User story 3 – Climate change**
English		
gdpr	consumer rights	ecology
privacy	consumer protection	green energy
data	civil rights	environment
security	human rights	biology
data privacy	immigration	ecosystems
data security	health care	environmental
computer security	women's rights	biodiversity
information security	rights of citizenship	pollution
data protections	civil liberties	global warming
corporate governance	constitutional rights	climate change
network security	property rights	deforestation
data management	freedom of speech	polluting
information technology	voting rights	greenhouse gas
data rights	freedom of religion	overgrazing
intellectual property	freedoms	ecological
Cyber law	freedoms of citizens	energy
data processing	free speech	agriculture
cyberterrorism	democracy	agricultural
cybercrime	privacy	farming
data store	constitution	agronomy
cybersecurity	the consequence of the brexit	the consequence of climate change
cyber	blockchain	EU environmental policy
cloud computing	consumer protection in E commerce	sustainability transition
privacy rights	consumer rights and protection	economic social cultural rights
data processing-regulation	citizen rights in EU	climate smart agriculture
personal information security	migration	protection of environment
open data portal	equality of rights in EU	green economy
blockchain	foreign policy	environmental policies
the right to reuse data	citizenship	climate change policy
data technology	single market in Europe	sustainable development goals
data protection	social law in EU	nature conservation
the impact of AI on data privacy	social inequality	European union environmental regulations
data collection	health inequalities in Europe	offshore installations
impact of AI on education	consumer law	climate environment
the importance of ML	European commission	environmental taxes
Data privacy regulations	Retail regulation	climate change policies
ai		renewable energy
		sustainability impact assessment

		European climate change programme
French		
sécurité de données	l'egalité des genres	climat
open data portal	l'inégalités des sexes	environnement
gouvernance des données	des disparités dans la société	changement climatique
blockchain	la loi sociale	effet de serre
utilisation des données	le droit du consommateur	bouleversement météorologique
machine learning	l'immigration	fonte des glaces
ml	le droit de l'homme	impact des déchêts
ai	le droit du citoyen	gaspillages
artificial intelligence	le droit des femmes	Réglementations écologiques
intelligence artificielle	la précarité	changement de climat
données privées	les lois étrangères	taxe environnementale
données sécurisées	l'impact du brexit sur l'UE	déforestation
collection de données	le droit d'expression	pollution
données	l'inégalité sociale	energie verte
ressource de données	les migrants	écologie
RGPD	la citoyenneté	reglementations environnementales
GDPR		protection de l'environnement
information confidentielle		énergie renouvelable
données confidentielles		biology
la sécurité de données		biosphère
la gouvernance des données		planète verte
l'intelligence artificielle		réchauffement climatique
l'utilisation des données		panneaux photovoltaique
protection de données		combat contre le réchauffement climatique

5.5 Patterns

The pattern is a template Utterance assigned to an Intent, which contains syntax to identify Entities in a pre-defined sentence form.

TABLE 16: PATTERNS LIST

English Pattern	French Pattern
[i (am|'m|m)]	[(voudrais|souhaiterais|cherche)]
[(have|receive|send|bring)]	[(à propos|discutant|parlant)]
[(want|would like)]	[(un|une)]
[(human rights|hr)]	[(ecommerce|e commerce)]
[(give|provide)]	[(de|à|a)]
[(looking|searching |researching)]	[(de|du)]
[(please|pls)]	[(écrit|publié|rédigé)]
[(E commerce|ecommerce |e-commerce)]	[(article|revue|rapport)]
[(a|an|any)]	[(svp|s'il te plait)]
[(look for|search for|try to find)]	[(mon|ton)]
[(commission|comission)]	[(ta|ton)]
[(do not|don't|dont)]	[(au sujet|concernant|qui parle)]
[(written|published)]	[(ue|union européenne|union europeenne)]
[(document|article|paper)]	[(gdpr|rgpd)]
[(topic|subject)]	[(veux|voudrais|souheterais)]
[(program|programme|programe)]	[(y a til|y a t il|y at-il)]
[(interesting|interested)]	[(de|par|du)]
[(technology|tech|technologies)]	[(document|article|papier)]
[(related to|deal with|in relation to|concerns)]	[(topic|suject)]
[(in|from|around)]	[(eu|europe|ue)]
[(english|en)]	[(femmes|femme)]
[(gdpr|rgpd)]	[(homme|hommes)]
[(research|researchs|researches)]	[(IA|Intelligence Artificielle)]
[(eu|europe|ue)]	[(ML|Machine learning)]
[(french|fr)]	
[(women|woman)]	
[(men|man)]	
[(can|could)]	
[(interesting|interested)]	
[(AI|Artificial Intelligence)]	
[(ML|machine learning)]	
[(does not|doesn't|doesn't)]	

5.6 Configurations

The intelligent virtual assistant OPIA is a set of many Azure services and software. To use the POC as a starting point for industrialization, it requires to understand and handle this specific environment. It is possible to import this configuration from Deloitte Azure portal tenant to a new OP Azure tenant as the Azure portal allows to transcribe all services and configurations from one Azure tenant to another. Furthermore, Azure also allow to move items and services from one tenant to another, or to a new subscription or resource group. Thus, it ensures that all the Azure configuration for the OPIA POC won't be lost for the industrialization. However, it is advised to understand how these services are gathered and linked to deploy the OPIA and optimize it for industrialization.

FIGURE 9: OPIA AZURE SERVICES CONFIGURATION AND ARCHITECTURE

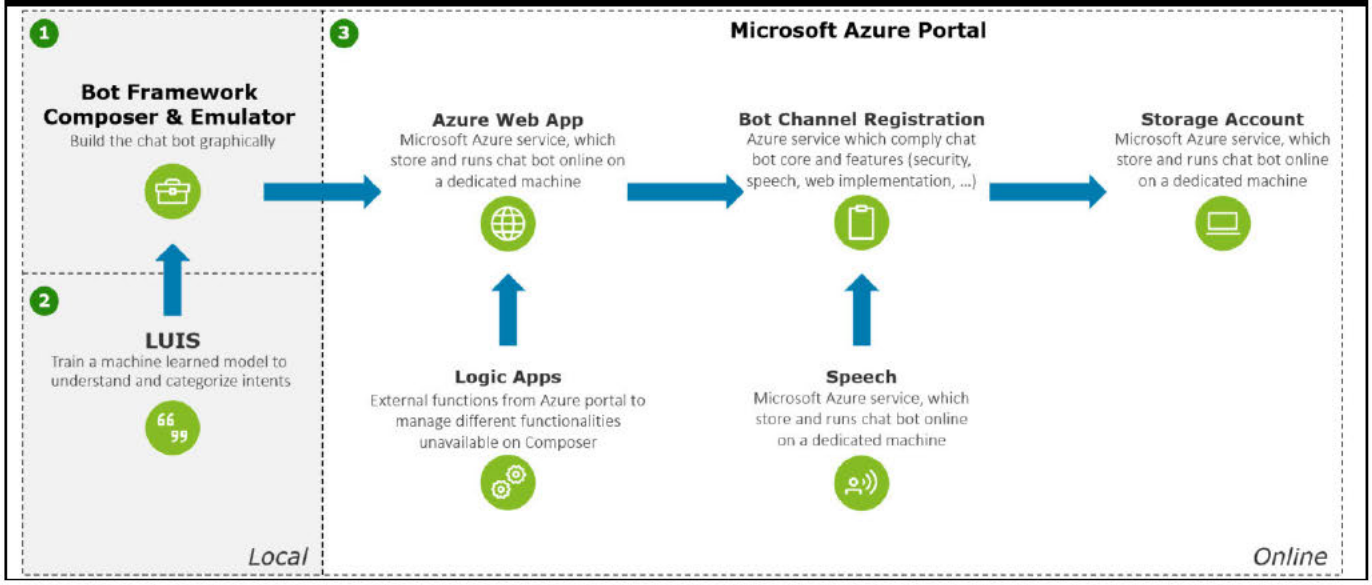


Figure 9 above shows the interaction between Azure software and services used in the OPIA POC. On the Azure portal, each service is gathered in one subscription and grouped by "Resource Group" (like a folder for Azure services). Each service is independent and linked to a specific subscription per service. For instance, to activate a speech feature, the speech service should be added to the resource group and linked to the other relevant services to produce the bot. However, if you want to gather and regroup these independent services into one domain name, Azure Front door or API Management will help.

The virtual assistant has been built with an open source software which is called "Bot Framework Composer". This software allows to build an intelligent virtual assistant using "Bot Framework" SDK from Microsoft with a graphic interface. This software is required to open and modify the virtual assistant developed by Deloitte.

TABLE 17: MICROSOFT BOT FRAMEWORK COMPOSER

Term	Description
Release	1.2.1
Extensions	None
Feature	Microsoft LUIS

To test and debug a virtual assistant from "Bot Framework Composer", the open source software "Bot Framework Emulator" is also required. It allows to launch and test locally without deploying the virtual assistant to Azure resource.

TABLE 18: MICROSOFT BOT FRAMEWORK EMULATOR	
Term	**Description**
Version	4.11.0
Extensions	None
Feature	Microsoft LUIS

In order to train the virtual assistant to recognize users' search queries, the virtual assistant built with "Bot Framework Composer" needs another service from Microsoft which is called "LUIS". LUIS is a cloud-based conversational AI service that applies Natural Language Processing algorithm to help the virtual assistant to recognize Intents from users' queries. To merge "Bot Framework Composer" with "LUIS" service, it is required to have an account on LUIS portal with a standard subscription. Once, this account is available, create what is called an "Authoring resource" on LUIS portal. An "Authoring resource" represents a folder where every LUIS model will be stored. Hence create a test model with "New App" and click on it. Go to "Manage" window, "Azure Resources" and then "Authoring Resource". Two values would be essential to merge virtual assistant from "Bot Framework Composer" with "LUIS" account: **Primary Key** and **Endpoint URL**. Finally, go to "Bot Framework Composer", on "Settings" window, add these values into the json file as below:

TABLE 19: LUIS CONFIGURATION ON BOT FRAMEWORK COMPOSER SETTINGS	
Term	**Description**
json	"luis": { "name": "name-of-your-LUIS-model", "authoringKey": "your-authoring-key", "endpointKey": "", "authoringRegion": "westeurope", "defaultLanguage": "en-us", "environment": "composer", "endpoint": " "authoringEndpoint": "your-authoring-endpoint-url"" },
name	It represents the name of your LUIS model on the LUIS portal when clicking on "Start the bot" or publishing the virtual assistant on "Bot Framework Composer".
authoringKey	The primary key value as explained above
endpointKey	Not required
authoringRegion	Region of LUIS account, here it is "westeurope"
defaultLanguage	Language of LUIS model. "en-US" for English virtual assistant and "fr-FR" for French virtual assistant
environment	(Optional) Represents a LUIS model sub-folder
endpoint	Not required
authoringEndpoint	Endpoint url value as explained above

Once, "Bot Framework Composer", "Bot Framework Emulator" and "LUIS" has been configured, it is now possible to launch the virtual assistant locally. However, to access the virtual assistant on OP Portal, it is required to publish it on Azure.

First step would be to publish the virtual assistant on an Azure cloud. Create an Azure account and a "Resource Group", which is like an Azure service folder. Then create an "App Service plan" resource which defines a set of compute resources for a web app to run.

TABLE 20: APP SERVICE PLAN	
Term	**Description**
Name	Luis-csharp-bot-opia
Location	West Europe
Subscription	S1
Operating System	Windows

Once, the App Service plan resource has been create, it is now possible to rent compute resources to store the virtual assistant on Azure cloud. Now, create a "Web App" resource to host the published virtual assistant.

TABLE 21: WEB APP SERVICE	
Term	**Description**
Name	Opia-app
Runtime stack	.NET Core 3.1
Region	West Europe
Subscription	S1

Creating a "Bot Channel Registration" resource is also required to use the virtual assistant on different channel.

TABLE 22: BOT CHANNEL REGISTRATION	
Term	**Description**
Bot handle	Bot-channel-opia
Location	West Europe
Pricing tier	S1
Application Insight	Off
Messaging Endpoint	None

Finally, a "Speech" service is also required to create the speech feature on your virtual assistant. It is advised to create two "Speech" services, one for English users and another for French users.

TABLE 23: SPEECH SERVICE	
Term	**Description**
Name	Speech-OPIA-POC-S0
Location	West Europe
Pricing tier	S0

Now that every resources has been created, they need to be configured to work together. First go to the "Bot Channel Registration" resource in "Settings" options. Click on "Enable Streaming Endpoint" and fill "Messaging Endpoint" with your "Web App" resource url with "api/messages" at the end. For instance, the url of the Opia-app "Web App" service which has been described above is "https://opia-app.azurewebsites.net", hence the "Bot Channel Registration" Bot-channel-opia "Messaging Endpoint" would be "https://opia-app.azurewebsites.net/api/messages". Furthermore, in the "Web App" resource,

go to “Configuration”, “General Settings” and tick “Web sockets” on “On” and “Save”. Now the Web App resource is connected to the “Bot Channel Registration” resource. Now, the “Bot Channel Registration” needs to create “Channels” to implement the virtual assistant on a website and add the “Speech” service. Go to “Bot Channel Registration Resource”, on “Channel” option, click on “Direct Line Channel” and create a new channel, do the same thing for “Direct Line Speech Channel” and select the “Speech” service in “Cognitive service account” (for instance “Speech-OPIA-POC-S0” as described below).

Now, the virtual assistant can be called from any website with the “Secret key” from “Direct Line” and “Speech Direct Line”. With these keys, it is also possible to manage users’ connections and provide token in web cookies to get access to the virtual assistant.

Only one part is remaining, actually publishing the virtual assistant from “Bot Framework Composer” to Azure cloud. To do so, go in “Bot Framework Composer”, in “Publish” window, “Add new profile” and select “Publish bot to Azure” then add a json file as below:

TABLE 24: JSON FILE TO PUBLISH THE VIRTUAL ASSISTANT TO AZURE

Term	Description
Json	{ "accessToken": "xxxxxxx", "name": "", "environment": "dev", "runtimeIdentifier": "win-x64", "settings": { "luis": { "authoringKey": "xxxxxx", "authoringEndpoint": "xxxxx”, "endpointKey": "", "endpoint": "”, "region": "westeurope" }, "MicrosoftAppId": "xxxxxx", "MicrosoftAppPassword": "xxxxxx" }, "hostname": "xxxxxx", "luisResource": "xxxxxxx” }
Access token	Token from Azure API. Download Azure CLI, type “az login”, connect to the Azure account and type “az account get-access-token”. The token is available for 1 hour
Environment	Same than LUIS configuration, represents a sub-folder for model in LUIS portal
LUIS	Same json than LUIS, explained above
MicrosoftAppId	The Id of the “Bot Channel Registration”, available in “Settings” in “Microsoft App ID” on Azure account
MicrosoftAppPassword	On “Microsoft App ID”, click on “Manage”, then “New client Secret”, a new password and will be displayed only once. Store it carefully and add the password here

hostname	Represents the name of the "Web App" where the virtual assistant will be published
luisResource	The name of the Authoring-resource on LUIS

Once, this json file has been filled, click on "Publish to selected profile", a "Success" message should appear. Now, the virtual assistant is on Azure cloud and accessible through "Bot Channel Registration". It can be called on OP Portal with Primary Key of "Direct Line" and Primary Key from "Direct Line Speech".

Another important part of the virtual assistant consists of using the OP Portal Search API. It is important to notice that "Bot Framework Composer" doesn't handle XML from REST API response. Hence, a "Logic app" services has been created to convert XML response to JSON response. An "Integration account" is needed to host a "Logic app" service. This is it is a secure, manageable and scalable container for the applications and maps.

TABLE 25: INTEGRATION ACCOUNT

Term	Description
Name	Integration-account
Location	West Europe
Pricing tier	Standard

TABLE 26: LOGIC APP

Term	Description
Logic app name	Integration-account
Region	West Europe
Integration service environment	Integration-account

5.7 OPIA revision history

TABLE 27: PROPOSED INCORPORATION OF COMMENTS		
Improvement areas	**Included in POC scope**	**Backlog for implementation**
Formatting & design	• Chatbot displays (blues, top bar, icons) compliant with OP brand specifications • To open documents, the whole area (title & icon) hyperlinked • Immediately get document proposed	• OPIA responses revised in terms of length • Consider visibility & placement of microphone • Document description (not grey on grey)
Speech	• Voice prompt explains the next steps to open documents and continue search more clearly • Speech Bot has the option to 'view all search results'	• Open a document using voice • Speech flow navigated through the keyboard
Functionality and flow	[ALL BOT VERSIONS] • EuroVoc codes translated to words [FRENCH SPEECH VERSION] • «Je cherche un document sur» replaces «je cherche » • "Année dernière" option was not working correctly (fixed) • Format types are now also listed in bot speech bubble • Sometimes types appear in the bot speech bubble as EuroVoc, even though buttons are correct (fixed) [FRENCH WRITTEN VERSION] • There are 4 date metadata but only 3 are written in the message(fixed)	[ALL BOT VERSIONS] • Article format should link to the immediate proposed format by OPIA [FRENCH SPEECH VERSION] • Add option to restart search by stating "Je voudrais commencer une nouvelle recherche"

TABLE 29: ADDITIONAL COMMENTS	
Comment	**Clarification**
[OPIA bots] Filtering should not have ability to select "None of the above" option as well as all of the above suggestions.	This issue has been noted to Microsoft during development. This is fixed in version 1.3 of adaptive cards (the OPIA runs on V1.2 as Bot Framework Composer is not yet compatible with V1.3).
[OPIA bots] There is a repetition of the question: "What are you looking for" when the "ArticleTheme" entity is not detected but the "Search" intent is detected	As expected. This is a work-around to allow the bot to pick up topics outside the 3 users stories.
[FRENCH WRITTEN VERSION] An error message in English appears and not French: "A connection attempt failed"	Normal as this is not our error but the system timeout for a Bot Framework error where the message is in English. Not something we can control as other messages that guides the user to search or give another input

Deloitte.