European Commission

Study on Interoperability of Public Institution Chatbots, Interoperability with Consumer Assistants, and Question Answering Implementation

PART 2: INTEROPERABILITY WITH CONSUMER ASSISTANTS

Completion tracker

Task 02 - Version history

Version	Main adjustments	Owner	Date
V0.1	First draft of 1.2 Current state of consumer assistants	Deloitte	04/10/2024
V0.2	First draft of 1.3 Key considerations for interoperability with existing consumer assistants	Deloitte	11/10/2024
V0.3	First draft of 1.5 Regulatory review	Deloitte	18/10/2024
V0.4	First draft for 1.4 Viable approaches	Deloitte	25/10/2024
V0.5	First draft for 1.6 Implementation framework	Deloitte	31/10/2024
V0.6	First draft of Abstract, Introduction & conclusion	Deloitte	08/11/2024
V0.7	Comments on first draft of study report	OP	15/11/2024
V0.8	Enhancement of draft and comments addressed (excl. abstract, introduction and conclusion)	Deloitte	15/11/2024
V0.9	Enhancement of draft and comments addressed (abstract, introduction)	Deloitte	22/11/2024
V10	Added Executive Summary & 1.3.2.5 CA search relevancy section. Address additional comments and add info from provider discussions	Deloitte	29/11/2024
V11	Enhanced section following CA meeting: 1.3.2.5	Deloitte	06/12/2024
V12	Deloitte addressed all comments & cleaned overall flow; addition of PoC findings & technical specifications	Deloitte	16/06/2025
V13	Added information on cost estimates for industrialization	Deloitte	09/07/2025

Contents

List of figures

List of tables

Executive Summary

Problem statement

The Publications Office (OP) has launched a project under the Digital Europe Program (DEP) with a focus on achieving interoperability between public institutions' chatbots and widely used consumer assistants (CAs). The primary objective is to improve the quality of information provided to citizens, ensuring it comes from trustworthy and reliable sources. At present, the information provided by CAs does not frequently originate from verified sources.

Key findings

The project has identified several key approaches to address this challenge:

- Direct Interoperability Solution:
  - Connecting public institutions' chatbots to CAs through the CAs services (e.g., Skills, Action) is proposed as a solution to offer citizens reliable information through the CAs they use in their daily lives.
  - To assess the feasibility of this approach, the OP demonstrated a connection between Amazon Alexa and Publio by creating an Amazon Alexa skill composed of various intents, which are voice control capabilities that allow Alexa to perform specific actions (e.g., Smart Home Control: “Alexa, turn on the lights in the living room” or “ Alexa, what is the weather in Luxembourg city tomorrow?”). This PoC will help understand how effectively a public institution's chatbot can provide trustworthy information through a widely used CA.
- Enhancing Website Indexing:
  - Public institutions can also enhance their websites to follow specific indexing standards. This approach will ensure that CAs can more easily identify and utilize reliable information from public institution websites.
  - Improved indexing will make it more likely that CAs will select webpages from public institutions, thus elevating the visibility of trustworthy information without the need of connecting the public institutions chatbot through the CA services.

Recommendations

Establishing interoperability between a public institution's chatbot and CAs will ensure reliable information is provided to citizens when they use their CAs to query the public institution's chatbot through a specific integrated skill set up to facilitate this interoperability. This approach leverages the existing capabilities of public institutions' chatbots, integrating them with widely used CAs without the need for developing entirely new systems.

Enhancing public institution websites with specific indexing standards will facilitate easier selection of these sites by CAs. This process will help CAs to prioritize pages from public institutions, thereby increasing the dissemination of trustworthy and verified information to the public.

Short conclusion

The proposed solutions capitalize on existing resources of public institutions to rectify the issue of unreliable information from CAs. By focusing on direct interoperability and/or improving website indexing standards, public institutions can significantly enhance the reach and impact of trustworthy information. These efforts will not require new solutions but will optimize the use of existing tools and platforms, thereby expanding the dissemination of reliable information through CAs. This approach ensures a broader reach and increased trust among citizens who already rely on CAs for information in their day-to-day activities.

Abstract

This study explores interoperability between public institution chatbots and widely available consumer assistants such as Alexa, Siri, and Google Assistant. Through this interoperable ecosystem, public institutions can more effectively meet citizen needs, modernize service delivery, and uphold principles of transparency and inclusiveness, all while providing trustworthy information to the public. It begins with an examination of the existing consumer assistants, comparing their capabilities and market availability. The discussion then extends to the benefits, challenges, and risks associated with consumer assistants. Key considerations for achieving interoperability are thoroughly explored, encompassing technological aspects, monitoring, contractual obligations, and security preconditions, as well as the influence of UX principles. Additionally, the study outlines viable approaches to consider — namely, the skill-based transfer approach and the gateway-based transfer approach. Regulatory considerations are also addressed, with a particular focus on the GDPR and the EU AI Act, and their impact on interoperability. Lastly, the study presents a comprehensive implementation framework designed to serve as a guide for public institutions or other potential users with chatbots, aiding them in developing an interoperable network through consumer assistants.

Glossary of terms

Term	Description
AI	Artificial Intelligence
ASK	Alexa Skills Kit
AVS	Alexa Voice Service
API	Application Programming Interface
ASR	Automatic Speech Recognition
CA	Consumer Assistant
DEP	Digital Europe Program
DoD	Definition of Done
DPIA	Data Protection Impact Assessment
GDPR	General Data Protection Regulation
HTTPS	Hypertext transfer protocol secure
IDE	Integrated Development Environments
IPRs	Intellectual Property Rights
KPIs	Key Performance Indicators
LLMs	Large Language Models
MAX	Multi-Agent Experience
ML	Machine Learning
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
OP	Publications Office
PoC	Proof of Concept
SDKs	Software Development Kits
SEO	Search Engine Optimization
SLA	Service Level Agreement
TTS	Text-to-Speech
UAT	User Acceptance Testing
UDC	Universal Device Command
UI	User Interface
UX	User Experience
VAs	Voice Assistants
VCS	Version Control System

1 Interoperability of public institution chatbots with consumer assistants

1.1.1 Introduction

This study aims to explore the potential for seamless interoperability between public institution chatbots and consumer assistants (CAs). With the increasing reliance on CAs for everyday tasks, ranging from setting reminders to accessing vital information, ensuring that public services information can be accessed through these commonly used channels becomes crucial. The significance of this interoperability lies in its ability to enhance both the accessibility and efficiency of public services. By facilitating the flow of trustworthy and timely information via preferred digital assistant platforms, public institutions can significantly improve the way citizens interact with public services. The study explores CAs and their capabilities, using selection criteria to pre-select CAs based on certain categories that would facilitate interoperability. From this selection the rest of the study will focus on five widely available CAs namely Alexa, Google Assistant (Gemini), Siri, Cortana (discontinued and replaced by CoPilot)[1], Bixby.

In this context, the primary objectives covered in this study are multifaceted. Firstly, it aims to achieve technical integration between public institution chatbots and CAs by exploring various technical solutions and deriving insights from potential approaches. This includes examining the architectures, protocols, and APIs needed to establish effective communication channels between these systems. Additionally, considerations were studied on how search relevancy is impacted to understand how to make trustworthy information from public institutions more visible for CAs. Secondly, the study seeks to give insights to ensure compliance with relevant regulations such as the artificial intelligence (AI) Act[2] and General Data Protection Regulation (GDPR)[3]. This is essential to maintain data privacy, security, and adherence to ethical standards while implementing these integrations. Lastly, the study focuses on improving user experience by offering reliable and accurate information to citizens through CAs. By leveraging these popular platforms, public institutions can provide a more user-friendly and intuitive way for citizens to access services and information.

The study also offers an implementation framework that serves as a structured roadmap for public institutions aiming to develop interoperability with CAs. This framework is subdivided into four main phases. The initiation phase involves establishing the project's scope, objectives, and stakeholders. It sets the foundation for the entire project by defining clear goals. The Proof of Concept (PoC) development phase focuses on creating and developing the interoperability framework and solutions. This involves conceptualizing how the integration will work, listing all necessary components for the development. Rigorous testing is conducted in the Testing phase to ensure reliability of the solution being developed. Finally, the deployment and monitoring phase entails implementing the solution in a live environment and continuously monitoring its performance and impact to confirm that the solution operates efficiently and meets user expectations.

The anticipated outcomes of this study are multi-dimensional, encompassing increased efficiency in public service delivery, enhanced user interaction and satisfaction, and expanded access to public services through widely used CAs, facilitated by the implementation framework serving as a guideline for public institutions to achieve interoperability with CAs. By achieving these outcomes, the project aims to demonstrate the feasibility and benefits of integrating public services with CA platforms. Ultimately, this integration has the potential to transform the overall user experience for citizens, making public services more accessible, efficient, and user-friendly. Through this interoperable ecosystem, public institutions can better meet the needs of their citizens, modernize service delivery, and uphold the principles of transparency and inclusiveness.

[1] Cortana is replaced by Copilot in the rest of the study as it was discontinued
[2] Regulation - EU - 2024/1689 - EN - EUR-Lex
[3] Regulation - 2016/679 - EN - gdpr - EUR-Lex

1.2 Current state of consumer assistants

The study focuses on Public institutions chatbots across the EU and their potential interoperability with existing popular CAs, which will be detailed in this chapter.

1.2.1 Distinction between agent and chatbots

A chatbot and an agent, while often used interchangeably, have distinct differences. A chatbot is typically designed for a specific task or set of tasks, operating through predefined scripts to interact with users via text or voice (Oyston, n.d.). It follows a linear path and is generally limited to answering questions, providing information, and executing simple commands.

In contrast, an agent, often referred to as a virtual or intelligent agent, possesses more advanced capabilities, leveraging AI to understand context, learn from interactions, and perform a broader range of complex tasks (Oyston, n.d.). Agents can engage in more dynamic and contextual conversations, adapt to user preferences, and integrate with multiple systems to offer a more personalized and seamless experience. Essentially, while chatbots are usually rule-based and deterministic, agents exhibit a higher degree of autonomy, learning, and decision-making, making them more versatile and sophisticated in handling user interactions.

In this study, the public institution chatbots are usually assumed to fall in the definition of chatbots, whereas CAs are usually more mature agents.

1.2.2 Overview of consumer assistants

AI voice assistants (VAs) or CAs, such as Amazon's Echo (Alexa), Apple's Siri, Microsoft's Copilot, and Google's Google Assistant, have profoundly transformed the way consumers interact with technology (McLean, Osei-Frimpong, & Barhorst, 2021). These sophisticated digital assistants leverage advanced natural language processing (NLP) and machine learning (ML) technologies to comprehend and respond to voice commands, facilitating a wide range of tasks from setting reminders and controlling smart home devices to providing weather updates and answering queries. Their integration into various devices, including smartphones, smart speakers, and even automobiles, has significantly streamlined daily routines, making information and services more accessible and convenient for users worldwide. As these technologies continue to evolve, they promise to further enhance personal productivity and enrich user experiences (UXs) through more intuitive and intelligent interactions.

1.2.2.1 Definition of consumer assistants

CAs are software agents operating on specialized speakers or smartphones. These agents continuously listen for an activation keyword, such as "Alexa" or "Hey Google" (Hoy, 2018). Upon detecting the keyword, the device captures the user's voice, interprets the language, and generates a response in real-time. CAs are capable of managing complex user requests and maintaining a dialogue (Alepis & Patsakis, 2017). The requirement for keyword activation allows users to remain in control of the usage of the system. For the remainder of this study, such voice- based activated systems will be considered as CAs in the context of how these can interoperate with EU public institutions' chatbots.

CAs first emerged as part of operating systems of phones, laptops, tablets (e.g., Siri) but then appeared as an application for different operating systems and smart speakers (e.g., Alexa). These assistants are embedded into multiple platforms, including:

Mobile phones: Available on iOS and Android devices, enabling on-the-go assistance.

Smart speakers and home systems: Integrated into devices like Amazon Echo, Google Home, and Apple Home Pod, allowing voice control of household environments such as diming the lights in the living room.

Laptops: Accessible via operating systems such as macOS.

Wearable tech: Found in smartwatches and fitness bands to assist with fitness tracking and notifications.

Automobiles: Embedded in car systems to provide navigation, entertainment, and hands-free communication.

Their integration into various operating systems enables them to offer a multitude of skills tailored to their specific domains (e.g., public institutions chatbots). These skills span a broad range of functionalities including:

- **Information retrieval:** Provide weather updates, news briefings, factual answers, EU policies, regulations, resources available, historical information.
- **Task management:** Set reminders, alarms, and calendar events (e.g., get an appointment at your local municipality).
- **Smart home control:** Manage smart lights, thermostats, and security systems.
- **Entertainment:** Play music, audiobooks, and podcasts or tune into livestream local events.
- **Communication:** Send messages, make phone calls, manage contacts, receive emergency alerts.
- **E-commerce:** Place orders, track shipments, and manage shopping lists.
- **Navigation:** Provide directions and traffic information in your city.
- **Services:** Get notifications from public transport services or information on educational services available.

Applying CAs to public institutions could lead to use cases such as providing information about local events, emergency alerts, and general information about city services (e.g., public transport services). For European Public Institutions, this can mean providing trustworthy and up-to-date and accurate information on EU policies, regulations, services, and resources available to citizens. Additionally, these assistants can offer more platform flexibility beyond the traditional computer and laptop setups commonly found in public institutions.

1.2.2.2 Selection of general-purpose consumer assistants to include in this study

A multitude of CAs are available today, each designed to cater to a variety of needs and preferences. Some assistants like Alexa, Siri, and Google Assistant, serve broad, general-purpose roles. Other assistants are tailored for specific tasks. This ranges over many disciplines, including:

- **AI work assistants**: These are designed to enhance the workplace productivity by managing tasks such as meetings, schedule organization, drafting emails, taking notes during meetings and/or producing key metrics during the call (e.g., who spoke the most, who asked the most questions). Examples include Fireflies.ai, AI Companion, etc (Wood, 2024).
- **AI Financial Assistants**: These offer personalized financial advice, assist with budgeting, and help users make more personalised investment decisions. Examples like Ally Financial, Cleo, and Tykr, provide comprehensive insights into spending patterns and financial health (Glover, 2024).
- **AI customer service assistants:** These assistants typically offer 24/7 support, proficiently answering frequently asked questions. Some existing examples like Front chat, Solve, and Fin can be designed to

address a broad spectrum of inquiries, extending beyond general customer support to cater to specific areas such as sales, marketing, and more.

- **AI Health Assistants:** These track fitness goals, remind users to take medications, and offer basic medical advice based on symptoms.
- **AI Shopping Assistants:** Such CA's help users discover products, compare prices, and manage shopping lists. In the educational sector.
- **AI Education Assistants:** These provide personalized learning experiences, assisting with homework and offering study materials.
- **AI Travel Assistants:** They plan itineraries, book accommodations, and provide real-time updates on travel conditions.
- **AI Legal Assistants**: CAs that can aid with legal research, document drafting, and answering legal questions.

While these assistants are part of the overall categorization as CA's, their support within their respective domains are valuable, but less relevant to this study as their functionalities are too specific and not directly aligned with our focus on interoperability for public institutions.

The following selection criteria will help us guide towards which, more general assistants, should be included to cater to broader public service needs.

Selection criteria

The criteria's below present the key categories that are required for a CA to be relevant in an interoperability network with a public service chatbot. These will be applied to select the scope of assistants to be included as part of the rest of the study:

- **SC1 - General purpose assistant:** A broad range of functionalities is key as assistants should not be limited to a specific domain. There should be the potential to integrate many skills smoothly to public institution chatbots compared to other more specialized assistants (e.g., e-commerce or customer service assistants).
- **SC2 - High overall user familiarity and consumer trust:** The CAs should have a large presence on the market. With many users already well-acquainted with these assistants, it will simplify the learning curve gentle and increasing the likelihood of higher adoption rates when public institutions chatbots interoperate.
- **SC3 – Advanced capabilities:** The level of technology inclusion in the CA's should include functionalities at the technological forefront (e.g., incorporate Large Language Models (LLMs) to enhance capabilities in understanding and generating human-like text). This selection criteria ensures that Public institutions select an interoperability partner that will be more likely to remain relevant in an ever-changing technology landscape.
- **SC4 - Embedded across various devices:** The CA should be able to function across a range of devices and services, making them versatile tools for connecting to public institution chatbots.
- **SC5 - Ecosystem integration:** The CA's provider should provide developer support, extensive resources, and Application Programming Interface (APIs) that facilitate the development and integration of new functionalities or services. These resources will simplify the setup of the interoperability network.

In this study, the focus is on general-purpose assistants based in the criteria above, which are highly available to consumers and integrated into a myriad of smart devices, enhancing their usability and accessibility (Klein, Hinderks, Rauschenberger, & Thomaschewski, 2020). The following general-purpose AI assistants have been selected based on the criteria for detailing in the study: Amazon Alexa, Apple Siri, Google Assistant, Microsoft Copilot and Samsung Bixby.

These CAs have become integral parts of daily life for millions of users worldwide. They are designed to handle a wide array of tasks, and this versatility makes them indispensable tools for both personal and professional use. One of the standout features of these assistants is their international recognition and widespread adoption, making

them familiar and easily accessible to a diverse user base. Additionally, they excel in cross-device functionality, seamlessly operating across various platforms such as smartphones, tablets, smart speakers, and even automobiles and often available in different languages. This multi-device compatibility ensures consistent UXs, allowing individuals to access their virtual assistants wherever they are, enhancing convenience. The following chapter will detail their functionalities and the specifics of each.

1.2.2.3 Market comparison of consumer assistants

Based on the above selection criteria, a few CAs have been chosen for further focus in the rest of the study. The scope is therefore exclusively on Siri (Apple), Copilot (Microsoft), Alexa (Amazon Echo), Google Assistant (Google), and Bixby (Samsung).

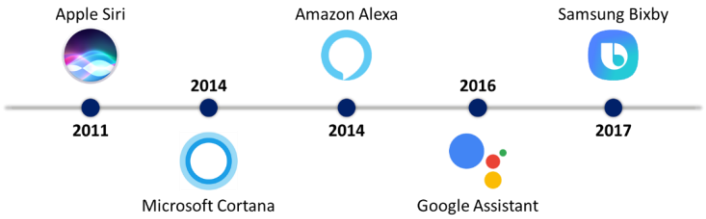


Figure 1. Overview of consumer assistants included in this study.

A. Siri by Apple

The market development of CAs took a significant leap with the introduction of Siri in 2011, widely regarded as the first AI personal assistant. Released as a standalone app in 2010, Siri was then integrated into the operating systems in 2011 of all Apple products, including the iPhone, iPad, Mac, Apple Watch, Home Pod, and AirPods (Hoy, 2018). It relies on voice recognition technology to respond to questions, make recommendations, send text messages, adding events to the calendar and perform various other tasks. Siri is also highly customizable, adapting its functionality to accommodate a user's language, searches, and preferences over time (Glover & Whitfield, 23 Popular AI Assistants, 2024).

Regarding its capabilities Siri has multi-user support voice recognition, meaning it can distinguish between the voices of different users when used on Home Pod but it does not provide extensive personalization options. It also supports more than 20 languages, including dialects, such as Arabic, Chinese, Danish, Dutch, English, Finnish, French, German, Hebrew, Italian, Japanese, Korean, Kannada, Telugu, Bengali, Marathi, Malay, Norwegian, Portuguese, Russian, Spanish, Swedish, Thai, Hindi, Turkish.

In 2016, Apple Inc. announced SiriSDK, a development kit that allowed programmers to integrate functions of their own apps as "Tasks" that Siri could perform. SiriSDK uses "Intents" as labels for user intentions and associates Intents with custom classes and properties (MobiDev, 2022). This capability could allow public institutions, to integrate their services and functionalities into Siri, enhancing accessibility and user interaction.

Siri is a cloud-based service that receives frequent updates to enhance its performance and capabilities. In 2024, Apple announced Apple Intelligence with new generative capabilities permitting richer language-understanding capabilities, Siri is more natural, more contextually relevant, and more personal, with the ability to simplify and accelerate everyday tasks. It can follow along if users stumble over words and maintain context from one request to the next (Apple, 2024). Additionally, users can type to Siri, and switch between text and voice to communicate with Siri in whatever way (Apple, 2024). They have also integrated ChatGPT allowing users access to its expertise. Users are asked before any questions are sent to ChatGPT (Apple, 2024).

B. Cortana by Microsoft (discontinued and replaced by BingChat and then CoPilot)[4]

Launched in 2014, this voice-activated assistant marked its debut as part of Windows Phone, evolving significantly over time. Initially designed for Windows Phone, the assistant eventually expanded to multiple platforms, becoming available on Windows 10, Windows 10 Mobile, iOS, Android, and Xbox.

The assistant offers a wide array of functionalities: from being a voice-activated assistant to managing schedules and reminders. It answers questions using Bing, provides weather updates, conducts web searches, and integrates seamlessly with Microsoft Office and third-party applications. The assistant supports multiple languages, including but not limited to English, Spanish, French, German, Italian, Chinese, and Japanese, making it accessible to a global audience.

After its initial launch with Windows Phone 8.1 in 2014, the assistant expanded its reach to Windows 10 desktops and mobile devices. It also integrated with Xbox and other Microsoft services. Over the years, its focus shifted towards enterprise productivity and integration with Microsoft 365 services. Consequently, its presence on mobile platforms scaled back as the emphasis moved primarily towards enterprise solutions. Ultimately, the service was discontinued in 2023, marking the end of its lifecycle within the consumer market (End of support for Cortana, 2024). Cortana has been replaced with new solutions such as voice access in windows 11 that helps you control a PC and write text using your voice or Microsoft Copilot an LLM that answers complex questions. With the discontinuation of Cortana, this study will no longer address it rather it will include its replacement Copilot.

In general, Copilot supports a wide range of languages for prompts, including Arabic, Catalan, Simplified and Traditional Chinese, Czech, Danish, Dutch, UK and US English, Finnish, Canadian and European French, German, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Polish, Brazilian and European Portuguese, Russian, Spanish, Mexican Spanish, Swedish, Thai, Turkish, Ukrainian, and Welsh.

Copilot is an advanced AI assistant that can enhance productivity from generating content for a user query to generating with a wide array of software tools from PowerPoint to Word. To being used for other tasks such as drafting emails, generating code, scheduling meetings, providing data insights, and even offering voice-driven interactions. Currently, its functionality is limited to devices such as laptops and phones.

C. Alexa by Amazon

Alexa by Amazon was introduced in 2014. Alexa is a cloud-based voice service that controls an entire smart device system. The engine behind Amazon Alexa runs on Amazon Web Services in the cloud, enabling Alexa to pick up user preferences and enhance its functionality over time.

Alexa responds to voice commands by providing information on a variety of topics, including products on Amazon, news, sports, weather, music, and more. It can manage tasks, set reminders, control smart home devices, and even make purchases. Alexa can also remember user voices and create individual preferences, making it efficient for personalized shopping, music preferences, calendar entries, etc. It supports multiple languages, including English, German, French, Italian, Spanish, Portuguese (Brazil), Japanese, and Hindi. Alexa is integrated into Amazon Echo devices and offers compatibility with third-party smart devices, such as security cameras, coffee makers, dishwashers, lights, and more (Koetsier, 2021).

The Alexa Skills Kit (ASK) is a software development framework that enables you to create content, capabilities, called skills, which provide users with an interactive voice interface to perform everyday tasks and control cloud-connected devices (What is the Alexa Skills Kit?, 2024). Alexa Skills are applications that brands can develop for Amazon Echo devices, enabling users to interact with Alexa through new functionalities. These "skills" allow Alexa to perform various tasks, such as arranging an Uber ride, ordering food, or controlling smart home devices. The skill

[4] Cortana is replaced by Copilot in the rest of the study as it was discontinued

might already be an existing Alexa pre-built model available or a custom one can be designed. Users can download these skills to expand Alexa's capabilities, thus continually enriching the UX. This potential could also extend to public institutions, allowing them to develop custom skills to provide citizens with accessible and trustworthy information and services, such as local government updates, public transportation schedules, and emergency alerts.

Amazon announced at the end of 2023 that they have optimized the Alexa LLM, making interactions more natural, expressive, and human-like (Rausch, 2023). Additionally, Alexa's conversation capabilities have been enhanced to cover any topic with customers (Alexa unveils new speech recognition, text-to-speech technologies, 2023).

D. Google Assistant

Google Assistant is an AI-powered virtual assistant developed by Google, designed to provide voice-activated, but also by text, assistance and control over a wide range of tasks and devices. Launched in 2016, it is integrated into Google's ecosystem and third-party devices, offering seamless functionality across various platforms such as smartphones, smart home devices, smart displays, watches, TV, car and more.

It assists users through NLP, enabling them to interact with their devices using voice commands. Google Assistant can perform a multitude of tasks, including:

- Providing information on news, weather, sports, and other topics
- Managing tasks such as setting reminders, alarms, and calendar events
- Controlling smart home devices like lights, thermostats, and security systems
- Sending texts, making calls, and navigating using Google Maps
- Facilitating hands-free purchases and reservations
- Playing music, podcasts, and videos
- Providing entertainment, jokes, interactive games

Google Assistant supports over 30 languages, including dialects, such as English, French, German, Spanish, Swedish, Dutch, Korean, Hindi, Finnish, Polish, Japanese and many more.

Recently, Google introduced the Gemini app for Android devices, replacing the Google Assistant app, while iOS users have transitioned to a revamped Google app (Wiggers, 2024). The Gemini-powered Google Assistant, currently exclusive to mobile platforms, supports a variety of input methods including images, text, and voice commands. Gemini, previously known as Bard, is an advanced LLM developed by Google. This generative AI exhibits an impressive ability to generalize and integrate various types of information such as text, code, audio, image, and video. However, it is important to note that the Gemini-powered Assistant utilizes Gemini Pro, rather than the more advanced and premium Gemini Ultra model, which is not available for free (Wiggers, 2024). More than 40 languages are currently supported, however for some of these languages, “Hey Google” voice commands and some Gemini action features are not supported.

E. Bixby by Samsung

Introduced in 2017 by Samsung Electronics, Bixby is a virtual assistant to better control Galaxy devices such as phone, watch, tv, buds and many more. From setting alarms for you, to changing tv channel or setting picture timer, calling, sending a message, starting your washer for you and so many more actions. Bixby is engineered to provide users with greater control and convenience, enabling a wide range of functionalities across Samsung devices including, but not limited to:

- Setting Alarms and Reminders: Bixby can set alarms, reminders, and calendar events
- Content Control: From changing TV channels to setting picture timers, Bixby offers hands-free control over entertainment devices and camera functions.
- Communication: Users can command Bixby to make calls, send messages, and read notifications.

- Smart Home Integration: With SmartThings integration, Bixby can start your washer, control lights, adjust thermostats, and manage other smart home devices.
- Information Access: Bixby provides weather updates, answers queries, offers restaurant suggestions, and provides news briefings.

Bixby supports the following languages: Chinese, English (US, UK, and India), French, German, Italian, Korean, Portuguese (Brazil), Spanish (Spain) and Spanish (Latin America).

Samsung has announced an upgrade to their CA, set to be released by the end of 2024 (Pessarlay, 2024). This new version will incorporate Samsung's own LLM, Gauss, aiming to significantly enhance its performance for mobile users (Kharpal, 2024).

Comparison summary

The market features several major players, with their own strengths, and market presence. The following table provides a comparison of these CAs based on various aspects, some of which were previously mentioned in 1.2.2.2~~1.2.1.2~~.

Table 1. Review of widely available consumer assistants' comparison

Feature	Siri	Copilot	Alexa	Google Assistant	Bixby
Company	**Apple**	**Microsoft**	**Amazon**	**Google**	**Samsung**
Launch Year	2011	2023	2014	2016	2017
Supported devices	iPhones, iPads, MacBooks, Apple Watch, Home Pod	Smartphones, laptops	Smart speakers (Echo), phones, tablets, laptops, smart TVs, and other smart home devices	Android phones, smart speakers (Google Home), tablets, laptops, smart TVs, and more	Samsung smartphones, tablets, smart TVs, and smart appliances (e.g., washing machines and fridges)
Market share (Thormundsson, 2023)	Dominates the mobile devices market	Recently entered the market	Dominates the smart speaker market	Healthy spread between all day-to-day devices (large market shar of mobile marked after Apple)	Behind in mobile devices, diverse presence in other devices.
Applications	Apple services (e.g., Music, Maps), broader Apple ecosystem	Microsoft services (e.g., Word, Excel, PowerPoint)	Amazon services (e.g., Prime, Music, Shopping), smart home devices (e.g., speakers, lights, cameras)	Google services (e.g., Search, Maps, YouTube), Android ecosystem	Samsung services, limited smart home appliances
Supported languages	20+ languages	20+ languages	8+ languages	30+ languages	8+ languages

Formatted Table

Feature	Siri	Copilot	Alexa	Google Assistant	Bixby
Written text input	Yes	Yes	Yes, on Alexa app	Yes	Yes
Third party integration	Moderate, primarily within Apple ecosystem	Moderate, limited to cross mobile & computer devices	High, extensive Skill library and third-party integration	High, works with many third-party apps and services	Limited, mostly Samsung-centric
LLM integration	Yes	Yes	Yes	Yes	Planned, end 2024

1.2.2.4 Consumer assistants' technology

CAs leverage advanced technologies to deliver seamless and intelligent interactions imitating real-life human conversations. At the core, **NLP** empowers these assistants to understand and interpret human language, making it possible to analyse user commands and queries. Complementing NLP is **ML**, which enables these systems to learn from each interaction, refining and enhancing their responses over time.

The following steps outline the process behind how these assistants operate (Zhou, 2023):

1. The assistant is activated with a keyword such as "Alexa", "Hey Siri", "Ok Google" making them ready to receive a command.
2. Voice recognition through **Automatic Speech Recognition (ASR)** technology is key, converting spoken language into text that can be processed by the system.
3. The NLP component, through **natural language understanding (NLU)** helps the system interpret the query and intent behind it.
4. Next, the system determines the correct answer based on the context. ML enhances this step by learning overtime the different variations of a query, and the most appropriate response.
5. Using NLP, specifically through **natural language generation (NLG)**, the focus shifts to generating a human-like response.
6. Finally, the system translates the text-based answer into a voice response with **text-to-speech (TTS)** software.

The immense computational power required for these tasks is provided by **Cloud Computing**, which ensures that data processing and response delivery are both swift and efficient. More recently, advancements in **Large Language Models (LLMs)** have further advanced the capabilities of these assistants, providing them with a more nuanced understanding and the ability to generate highly coherent, contextually relevant responses.

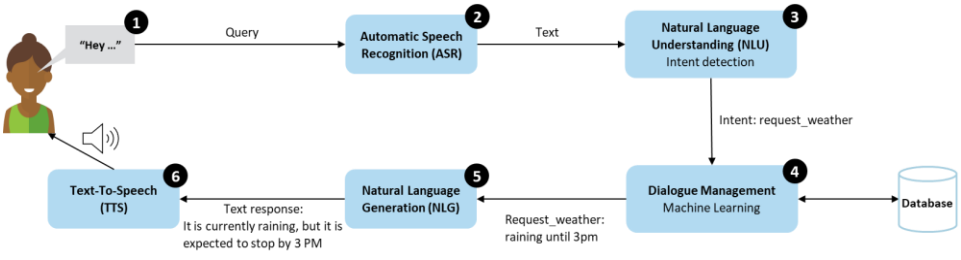


Figure 2. Consumer assistants flow overview

1.2.3 Benefits, Challenges & Risks

To comprehend the current landscape of CA and their services, the following section outlines the potential benefits, challenges, and risks that should be considered when utilizing these technologies.

1.2.3.1 Benefits

Benefits are defined as the advantages or positive outcomes that result from using CAs. By identifying and maximizing these benefits, individuals, organizations, or societies can create value, improve performance, and achieve desired objectives. Some key benefits of CAs that can benefit Public Institutions through interoperability are:

Table 2. Overview of potential benefits of consumer assistants

Benefit	Description	Public Institutions Use Cases
Convenience	As voice input is faster than keyboard input will enable consumers to spend less time to perform these actions (Budzinski, Noskova, & Zhang, 2019). This simplifies daily tasks by enabling hands-free operation. This functionality allows users to interact with the CAs without interrupting their ongoing activities, thereby promoting effective multi-tasking.	Integrate voice-activated visitor assistance at tourism offices.
Accessibility	Interacting with a CA requires no special knowledge or skills, just conversational ability. This ease of use makes it accessible to children, the elderly, and people with disabilities, enabling them to access various services more easily (Yaghoubzadeh, Kramer, Pitsch, & Kopp, 2013)	Use in public health services to assist elderly and persons with disabilities to book appointments.
Personalization	Learns user preferences and behaviors to deliver increasingly personalized and relevant responses.	Public libraries could offer personalized book recommendations based on current availabilities and user preferences.
Device coverage/ simple integration	Seamlessly connects with various devices and platforms, creating a unified UX across different environments (e.g., phone, computer, speaker).	Provide public emergency messages to any systems.
Routine management	CAs can help manage daily routines by sending reminders, setting alarms or meetings in the calendars.	Waste removal reminders can be set for different neighbourhood services
Language translation	Some CAs provide the ability to translate languages in real-time, useful for travellers or people learning new languages.	Use in immigration offices to assist non-native speakers.
Human-like interaction	CAs integration with LLMs enhances their capabilities, refining their current functionalities and introducing new features. This advanced integration enables more complex and contextual interactions.	For citizens to ask public health questions such as having a summarized answer on the new health guidelines for Covid-19 vaccinations.

1.2.3.2 Challenges

Challenges are defined as obstacles or difficulties linked to the use of CAs. They often arise when encountering constraints or uncertainties and can be addressed through strategic planning, innovation, or problem-solving skills. Some of the potential challenges that can hinder Public Institutions when connecting through interoperability to such devices include:

Table 3. Overview of potential challenges of consumer assistants

Challenges	Description	Mitigation actions
Accuracy	Understanding diverse accents, dialects, and speech patterns can be difficult, or background noise, leading to misinterpretations.	Public institutions can select a vendor that integrated LLMs and continually updates their models for regional accents recognition. Implement noise-cancellation technologies in public service areas.
Interoperability Issues	Different CA ecosystems (e.g., Alexa, Siri, Google Assistant) may not always work together seamlessly, creating compatibility issues with various devices and services.	Promote the use of open APIs and protocols for better interoperability across different platforms used in public institutions.
Data privacy and Security	Handling sensitive public data through CAs may raise privacy and security concerns.	Implement stringent data privacy measures and encryption. Restrict access to critical information based on user roles and ensure secure data storage and transmission practices.
User training and adoption	Public institution staff and users may need training to effectively use CAs, potentially requiring significant time and resources.	Provide comprehensive training programs and resources for staff. Develop user-friendly guides and FAQs.
Reliability and Dependence	Over-reliance on CAs may lead to operational issues if the technology fails or experiences downtime.	Establish fallback mechanisms, such as manual processes or alternative support systems, to ensure continuity of services. Regularly evaluate and monitor system performance to preempt and address potential issues promptly.
Integration Overhead	Integrating CA capabilities into various products and services could require development and maintenance efforts, which can be resource-intensive for institutions.	Employ modular and scalable architecture and leverage third-party tools and platforms to reduce development and maintenance efforts. Establish a dedicated team or collaboration with specialized vendors to streamline the integration process.

1.2.3.3 Risks

The use of CAs through interoperability with Public institution chatbots may pose risks, highlighting potential negative outcomes linked with uncertainty. Effective risk management includes identifying, analysing, assessing, and mitigating these risks to safeguard assets, reputation, and stakeholders. The risks identified include:

Table 4. Overview of potential risks of consumer assistants

Risks	Description	Mitigation actions
Privacy	Feature such as always-listening devices can raise concerns about data security and privacy. User can also worry about their conversation being recorded.	Add a rule to provide transparency to users on conversation recording and guidelines on their data collection and usage. The voice activation feature can also be configured to be disabled or to work with activation button (e.g., press the home button and then start speaking).
Data use	The way user data is collected, stored, and used can raise concerns, particularly around consent and data protection.	Implement clear data usage policies, anonymize data when possible, and provide users with easy ways to control and delete their data.
User Adaptation and Trust	Some users may find it difficult to adapt to or trust CAs, especially if they are not tech-savvy or have concerns about technology reliance.	Provide user-friendly tutorials, engage in trust-building initiatives such as transparency in AI decision-making, and offer robust support.
Carbon footprint	The use of AI tools, particularly those that demand substantial computational power, can result in elevated energy consumption and a larger carbon footprint. This issue is especially pertinent, as EU institutions are expected to lead by example in climate action and sustainability.	Implement energy-efficient models to reduce computational demands, use optimized data transfer protocols to minimize data processing load, select a more environmentally friendly LLM and CA, conduct continuous assessment and monitoring of the energy consumption and carbon footprint to mitigate the impacts.
Obsolescence by LLMs	CAs could be superseded by LLMs as they are more advanced and may integrate CA functionalities, providing more robust, versatile, and context-aware interactions.	Public institutions could select a virtual assistant that has already integrated LLMs in their solution.

CAs present a promising technology with the potential to greatly enhance convenience and efficiency in daily tasks and operations of public institutions when integrated though interoperability to their chatbots. They offer numerous benefits, including hands-free interaction and seamless automation, which can significantly improve UXs. However, the implementation and use of CAs are not without their challenges and risks. Addressing obstacles such as understanding diverse speech patterns, along with mitigating risks related to privacy, security, and bias, are crucial for maximizing their positive impact. Overall, while CAs hold immense potential, careful consideration and proactive management of their associated challenges and risks are essential for their successful and responsible adoption or integration especially with the emergence of LLMs.

1.3 Key considerations for interoperability with existing consumer assistants

Public service chatbot are revolutionizing the way citizens interact with government services by providing instant, accessible, and efficient responses to their queries and needs. However, to maximize their reach and usability, it is important to consider their integration to widely available consumer CAs such as Amazon Alexa, Apple's Siri, Google Assistant, and Samsung's Bixby.

Achieving interoperability, the ability of different systems to work together seamlessly, between CAs and public chatbots involves several considerations. The following section will present some key aspects necessary to consider the foundation of interoperability.

1.3.1 Basic interoperability preconditions

Integrating public service chatbots with widely available CAs requires meeting or considering several essential preconditions. These can include ensuring API compatibility, obtaining necessary licenses, and complying with relevant data privacy and security regulations.

1.3.1.1 Technology considerations

Successfully integrating public service chatbots with widely available CAs like Siri, Alexa, Google Assistant, and Bixby relies heavily on achieving integration compatibility with each platform's unique APIs and Software Development Kits (SDKs). The use of SDKs and API connections, enables developers to integrate and augment their functionalities.

An SDK is a set of tools, guidelines, and programs used by developers to create applications for specific platforms, while an API is a set of protocols and definitions that allow different software applications to communicate with each other. Many of the SDKs typically have APIs integrated as part of the toolbox. Some platforms also include Integrated Development Environments (IDE's) that provides comprehensive facilities to computer programmers for software development. Below, the study details the technical solutions from each platform to foster interoperability between CA's and public sector chatbots.

In the following comparison, systems like Copilot have not been included because it is primarily a professional Microsoft productivity tool based on LLMs, rather than a consumer-focused Voice Assistant (VA).

Table 5. Tools for integration with consumer assistants

System	Description	Basic steps	Documentation
Siri	**SiriKit:** SiriKit enables your iOS and watchOS apps to work with Siri. It allows you to handle voice commands and perform related tasks within your app.	1. **Prepare App and configure provisioning profile:** Enable Siri in Xcode by adding the "Siri" capability. 2. **Define Intents:** Create an Intents Definition File and define the intents you want to support by adding "Intent" and "Intent Response" classes. Configure parameters for each intent. 3. **Handle Intents:** Create Intent Handlers and implement classes that conform to the appropriate protocols (INExtension, INIntentHandlerProviding) to handle the intents. 4. **Testing:** Use Siri or the Shortcuts app to test your intents. Debug using Xcode and the Console.	SiriKit | Apple Developer Documentation

System	Description	Basic steps	Documentation
		5. **Design Custom Responses** Set up custom user interface (UI) for Siri (using UI extension) if required.	
	Siri shortcut: Siri shortcuts are a feature introduced by Apple that allows users app to perform tasks or trigger actions quickly using Siri voice commands. They provide a way to automate tasks in iOS apps and enable a seamless interaction experience for users. Connecting a chatbot to Siri would require simple tasks such as creating a shortcut (user activity or using SiriKit intents) and setting an invocation name (through Siri Shortcut apps).	1. **Create a Siri shortcut**: Define an actionable shortcut within your iOS app that represents an interaction with your chatbot. This involves creating a user activity or using SiriKit intents. 2. **Set an invocation name:** Rename the shortcut to specify a custom voice command, or invocation name, that users will use to call the chatbot via Siri. This can be done through the Shortcuts app. 3. **Connect via API Key:** Ensure your chatbot can communicate with your iOS app by integrating an API key. This key will authenticate and authorize requests made from the iOS app to your chatbot's backend.	Shortcuts User Guide - Apple Support
Alexa	**Alexa Skills Kit (ASK):** To integrate public service chatbots with Amazon Alexa, developers could leverage the ASK. This is a collection of self-service APIs, tools, and documentation that allows developers to create custom Skills for Alexa-enabled devices. Developers can use ASK to build tailored Skills. These Skills could be tailored to offer a specific public institutions service such as appointment bookings, information retrieval, and notifications. *A prerequisite would be to have an Amazon developer account* (Connect a bot to Alexa, 2022).	1. **Create an Alexa Skill**: Create a custom skill, define an invocation name and intent and slots to capture user interactions. 2. **Backend Endpoint:** Connect via the Hypertext transfer protocol secure (HTTPS) service endpoint by adding your API endpoint. 3. **Test**: Test your skill to ensure that it correctly handles requests and responses. 4. **Deploy and certification**: Deploy and submit your	Create Alexa Skills Kit | Amazon Alexa Voice Development • Understand Custom Skills | Alexa Skills Kit • Test and Debug Your Skill | Alexa Skills Kit • Certification Requirements for Alexa Skills | Alexa Skills Kit

System	Description	Basic steps	Documentation
		Alexa skill for certification by following the certification requirements.	
	Multi-Agent Experience (MAX) toolkit – a framework designed to foster interoperability and provide a unified UX among multiple voice agents on a single device. By leveraging the MAX toolkit, developers can create sophisticated multi-agent applications that harness the unique capabilities of various voice services, resulting in smoother and more integrated interactions for users on their devices. They have already established interoperability with other products such as a Garmin GPS, and have provided demos with many others like a LG Smart TV, a Samsung appliance, the Korean SK telecom speaking agent, etc.[5] *Note: According to Amazon Alexa support team, this solution is no longer supported*	Refer to the GitHub of the Toolkit for all information and requirements.	Multi-Agent Experience (MAX) Toolkit Multi-Agent Design Guide (amazon.com)
Google Assistant	**Actions SDK:** Public institutions can achieve effective integration with Google Assistant by utilizing the Actions on Google platform and Dialog flow to create conversational agents and dynamic interfaces for public service chatbots. Google assistant uses the Actions to enable developers to create conversational agents with a knowledge base and making it available for Google assistant to use. *Note: the conversational actions have been discontinued on June 13, 2023* (Conversational Actions, 2024).	1. **Create an Action:** Define an invocation name for user to use. 2. **Develop interaction model:** Define the intents and provide training phrases to help Google Assistant understand user request as well as entities. 3. **Set up dialog flow:** create intents in Dialog flow for user actions and desired interactions. 4. **Endpoint:** Set up a webhook, an HTTP request, to handle requests from dialog flow. This webhook will process user input and communicate with your chatbot API to return responses to Dialog flow.	Overview | Google Assistant SDK | Google for Developers

[5] Multi-Agent Products (amazon.com)

System	Description	Basic steps	Documentation
		5. **Test and deploy:** Use the Dialog flow console to test the Action. Deploy and submit the Action for review and certification by Google.	
	App Actions: This capability enables users to access an Android app's feature by voice. *A prerequisite would be to have an android app of the public institution chatbot on Android*	1. **Identify in-app functionalities for voice trigger:** Access Google Play Console and identify functionalities in your Android app that users might want to jump to with a spoken request. 2. **Implement intent:** Review built-in intent to match key functionality and user flow in the app. Custom intent can be used to extend your app with App Actions. 3. **Test:** test the App Actions for your app through a preview by triggering queries by voice. Additionally, test your App Actions with additional testes through test release of your app. 4. **Request review and deployment:** Your App Actions needs to be submitted for approval before being published.	Google Assistant for Android | Android Developers App Actions | Google for Developers How to use Gems, Google's custom AI tools
Bixby	**Bixby Developer Studio IDE:** An IDE for developing Bixby Capsules, that includes simulators, debugging tools, and other utilities. Integrating public institution chatbots with Samsung's Bixby would involve leveraging Bixby Capsules and utilizing the Bixby Developer Studio. Capsules are modular applications that extend the capabilities of Bixby. These Capsules enable users to perform tasks using natural language commands to connect to apps such as Google Maps,	1. **Create a Capsule:** use the Bixby Developer to create a Capsule project. 2. **Define interaction model:** Add Actions, determining what Bixby can do, and Concepts, explain what Bixby knows. 3. **Develop the Capsule:** Use JavaScript or Bixby's modelling language to code the logic for your Capsule. This includes	Bixby Developer Center (bixbydevelopers.com)

System	Description	Basic steps	Documentation
	Gmail to Bixby-enabled devices (Overview, n.d.).	handling user interactions and integrating with your chatbot's API by setting up endpoints. 4. **Test the Capsule:** Use the built-in testing tools to simulate user interactions. 5. **Deploy and publish:** Submit the Capsule for review, once approved it will be available for users to enable.	

The table above provided insights on the steps to be followed, ready for download by users on the appropriate platforms. However, given that consumer VAs like Google Assistant, Alexa, Siri and Bixby are currently undergoing rapid evolution, by integrating LLMs, the integration methods and best practices described above may soon become outdated. As these AI-powered technologies continue to improve their understanding and interaction capabilities, significant changes and new features that could simplify or alter the way developers connect chatbots and other services to these platforms could be expected.

1.3.1.2 Monitoring considerations

Effective monitoring and analytics are crucial for ensuring the performance, reliability, and continuous improvement of your chatbot when integrated with CAs like Alexa, Siri, and Google Assistant. Each platform provides tools and features that enable developers to track usage patterns, identify issues, and gather user feedback. Monitoring plug-ins available in vendors, include:

- **SiriKit:** No explicit monitoring functions through analytics or event logging services. Instead, SiriKit provides functionalities to handle user interactions through defined intent handling support or through donating certain user activity and intent handling to get proactive suggestions from Siri[6].
- **Alexa Skills:** ASK Analytics[7] provides detailed analytics for your Alexa Skills within the Alexa Developer Console. You can monitor key metrics such as the number of unique users, the number of sessions, utterance count, and interaction paths. These insights help you understand how users are engaging with your skill and identify any issues or opportunities for enhancement.
- **Google Actions:** Actions on Google Console[8] offers comprehensive analytics for Actions on Google. Key metrics include usage trends, user engagement, session length, retention rates, and drop-off points. These metrics provide valuable insights into user interactions and the overall success of your action.
- **Bixby Developer Studio IDE:** Bixby Developer Studio offers a comprehensive set of tools for monitoring, debugging, and optimizing the performance of your Capsule including analytics to monitor usage and performance metrics, logging to access detailed interaction logs, tests and debugging, analysis on NLU performance and accuracy, error Handling, etc[9].

[6] INInteraction | Apple Developer Documentation
[7] Monitor Your Skill Metrics and Earnings | Alexa Skills Kit (amazon.com), View Skill Metrics | Alexa Skills Kit
[8] Analytics | Actions console | Google for Developers
[9] Using Bixby Developer Studio | Bixby Developer Center

Public institutions should consider what type of information they would like to track as part of monitoring the interoperability between their Chatbot's skills that are linked to CAs.

1.3.1.3 Licensing and contract considerations

In the effective implementation of interoperability between systems, several contractual aspects need to be considered. The following outlines some key documentation to potentially address. Always confirm with the institution's legal department before any contracts are made for alignment with internal policies:

A. Licensing considerations
   - **Licensing for Third-Party Integrations**:
     a. API Usage Licenses: If the chatbot integrates third-party APIs or services, ensure that appropriate licenses are obtained for these integrations.
     b. Open-Source Components: Verify compatibility and compliance with licenses of any open-source components used in the development of the chatbot.

B. Data Privacy
   - **Data Processing Agreement:** Establishes rights and obligations of contractual parties for handling shared personal data during the interoperability process and determines details regarding processing of personal data as required by GDPR.
   - **Data Protection & Privacy:** Ensures GDPR compliance in data collection, processing, use, and sharing; addresses GDPR principles and security measures.
   - **Privacy Notice:** A privacy notice is a statement including details about processing of individuals' personal data conducted by a business. It is a requirement under GDPR to fulfil transparency principle.

C. Other terms
   - **Service Level Agreement (SLA):** This defines the level of service expected by the customer from a supplier, laying out the metrics by which that service is measured, and the remedies or penalties, if any, should the agreed levels not be achieved.
   - **Intellectual Property Rights (IPRs):** The EIF tends to favour the use of open standards and specifications regarding IPRs and has been updated in 2017 in hopes to further facilitate interoperability.
   - **Copyright:** The EU Software Directive provides some allowances with copyright law, stating reverse engineering to achieve interoperability is not an infringement, given specific conditions are met (European Parliament & Council, 2009).

Having a clearer understanding of the contractual requirements to consider for interoperability with CAs, the focus now shifts to the security considerations.

1.3.1.4 Security considerations

When enabling interoperability between a public chatbot and CAs such as Alexa, Siri, or Google Assistant, security considerations must be addressed.

Interoperable chatbots' security can be fortified via specific measures. Key aspects include exclusive communication with desired assistants and abuse protection by employing data encryption, authentication, and protocols.

Ensuring communication with only desired assistants

Exclusive communication with desired chatbot is already ensured as the public institution chatbot would need to publish on the CA, therefore ensuring only chatbots published are available. Moreover, communication between the desired CA and public institution chatbot is done when a wake word, invocation word or intent is recognised to redirect to the corresponding bot. A limitation, mentioned in section 1.3.2.4, would be for the CA to differentiate which chatbot is triggered when multiple wake words are mentioned in one query.

Protecting assistants and users

Preventing resource abuse: Rate limiting prevents resource abuse by restricting request rates from each CA and public institution chatbot, ensuring system performance. Additionally, authentication and authorization mechanisms limit interaction to authorized chatbots. Protecting chatbots from abuse can be done using HTTPS encryption and tokens. HTTPS encryption, a widely used protocol, secures assistant communication by encrypting transmitted data, preventing unauthorized access. In interoperability, this safeguards confidentiality and integrity of shared information. A Token or Key system enhances security by assigning unique identifiers to authorized CA or chatbot, controlling interactions and reducing unauthorized access risk.

It is important to understand that the encryption of data by a public institution chatbot does not guarantee that the data will remain encrypted once it enters the ecosystem of a CA. To mitigate this risk, a contractual agreement could be envisioned. This agreement would require the CA to adopt and adhere to the public institutions security policies regarding encryption and data handling for interoperability. Such measures would ensure that the security protocols are maintained consistently. Note, that potentially the CA internal policies might already cover this aspect.

Regulatory compliance: CAs operating in Europe are required to comply with regional regulations such as the GDPR or the EU AI Act, refer to section 1.5 Regulatory outlook. This ensures that personal data is handled in accordance with stringent privacy and security standards, further enhancing the protection of interoperable communications.

Certification and review processes: Prior to publication, a skill, action, or shortcut typically undergoes a review process, ensuring that it meets necessary security standards and checks. According to AWS, this process usually takes a few days.

1.3.2 The impact of UX principles on interoperability

In terms of UX/UI considerations for interoperability with CAs, there are several factors to keep in mind.

1.3.2.1 Skill accessibility and user awareness

As outlined in 1.3.1.1, to make the public institution's chatbot available on CA platforms via a skill, it must go through a review process specific to each platform. Upon successful validation, users will be able to see the skill listed and will need to invoke it to interact with it on their devices. This review ensures the chatbot meets the platform's standards for quality and security, making it trustworthy for end-users.

It is still important for users to be aware of the capabilities of the CAs and, in some cases, proactively download and enable the appropriate skills. Here are several strategies to enhance user awareness and interaction, by promoting the availability of the skill:

- **Visibility through Public Institution:** Public institutions should prominently feature information about the available chatbot skills on their official websites and other digital platforms. This includes providing details on how to enable and use these skills. It is also possible to share the skill through different platforms such as via email, text, or social media (e.g., Facebook)[10]. It is recommended for the public institution to promote through their own administration channels utterances and use cases.
- **Visibility through CA:** Utilize the CA's application, such as the Alexa app, to promote the availability of the chatbot skill.

1.3.2.2 Skill relevance

To ensure that a skill is relevant and remains so, the Public institution will need to consider the following aspects:

[10] Alexa Skill Blueprints (amazon.com)

- **Content relevance:** Ensure that the information provided by the chatbot is relevant and up to date. Public institutions can edit their skills to make updates available to users promptly. If a skill needs to be discontinued, it can be disabled (making it unusable but not deleted) or deleted entirely.
- **Localization:** It can be specified in which countries the public institution would like to share the skills in or select all the countries where the CA distributes them. This is key for specific public institution chatbot skill that are linked to a specific region, municipality or city (e.g., reminder for waste removal services from your local municipality is only relevant if the required localization is set up well). Ensuring that users connect to the correct region and localisation of the skills are part of the installation process.
- **Guidance on usability:** Public institutions can provide detailed tutorials and step-by-step guides on their website. These guidelines should cover how to enable the chatbot skill (e.g., how to enable and what the invocation name is), and examples of common commands. This could include a short instructional video and demo.

One main benefit of integrating Public institution skills with CAs are to ensure that the most relevant information and data sources are made available to end users. Such effective content delivery is essential for a positive UX.

1.3.2.3 Communication aspects

The following communication strategies needs to be considered:

- **Languages:** A skill can support a single language or multiple languages that are available on the CA. The consumers that interact with the CA in a particular language can then use the skill that support that language. It should be noted that if a user sets their CA device in French, they might only be able to enable and use the skills supported in French. In case these users want to import skills in an unsupported language, the device language might need to be changed[11].
- **CA tone:** Adopt a tone that suits the nature of the public sector, keeping in mind that while the content is provided by the public institution, it will be delivered using the CA's voice. Additionally, be aware that users have the option to change the CA's voice, which may affect how the content is perceived.
- **Interaction channels**
    - **Voice channel:** Voice interactions could take two distinct approaches.
        - **A.** The CA's voice remains, even when the public institution interacts through the CA's interface (What is the Alexa Skills Kit?, 2024). This creates a unified auditory experience for users, as for example they will hear the familiar voice of Alexa. This approach emphasizes seamless transitions and may leverage the CA's established tone, and language preferences, to maintain coherence however, there might be slight changes to the personality as the conversational style of the public institutions chatbot will remain.
        - **B.** The second approach would involve differentiating the voices when the public institution's chatbot interacts. For instance, this means users would hear both Alexa's voice, and a different distinct voice representing the public institution that would reply when specific queries or invocation words are used. This can enhance clarity and context for users as they recognize shifts in the source of information and what kind of queries they could ask.
    - **Text channel:** For the text channel, interoperability means displaying textual content on the CA's screen, if available, when interacting with the public institution's chatbot. This is not a frequently used feature and is also not available across all devices as a screen would be required. This could include not only standard text but also rich text elements like hyperlinks, images, and formatted documents. Some CAs and their devices require the user to enable the typing feature in their settings.

[11] how to change the bixby voice language | Samsung LEVANT, Ask Alexa to Speak in Multiple Languages - Amazon Customer Service, Develop Skills in Multiple Languages | Alexa Skills Kit , How can I make Siri understand two language... - Apple Community, Change the language of Google Assistant - Android - Google Nest Help

Table 6. Text ability across the CAs

	Siri	Alexa	Google Assistant	Bixby
How to use text input	Use typing to Siri	Use the Alexa app to type commands	Google Assistant app allows text input	Bixby app allows text commands
Availability on Devices	Available on iPhone, iPad, Mac	Available on mobile devices via the app	Available on Android, iOS, and smart devices	Available on Samsung devices
Feature Activation required	Yes, needs to be enabled in settings[12]	No, text activation is enabled by default[13]	No, it is enabled by default	No, it is enabled by default

Textual responses need to be clear, concise, and might need to be formatted appropriately to fit the display characteristics of various CA devices. Customizing text content for screen size, resolution, and context ensures clarity and ease of reading. For instance, when a user asks about renewing a driver's license, the response text can be accompanied by relevant hyperlinks and a clear, bulleted list of steps displayed aesthetically on the CA's screen. To summarize if users prefer to input queries via text, they can do so seamlessly within the dedicated apps for CAs like Alexa and Google Assistant, as this feature is enabled by default. For Alexa, the text input option is readily available within the Alexa app. Similarly, users can access this feature in the Google Assistant app without any additional setup. On the other hand, for Siri, users must first manually enable the text input capability within the device settings to utilize text-based interactions. This varied approach across different platforms offers flexibility but requires users to be aware of the specific enabling procedures for each CA.

Effective communication strategies for CAs connected to public sector chatbots must consider language support and chatbot tone and address compatibility to ensure good UXs. These considerations will enhance user engagement and provide a more seamless and effective experience.

1.3.2.4 Conversational flow elements

Interoperability of Public institution chatbots with CAs, may lead to certain limitations in customization of the skill linked to the CA, particularly in terms of conversational flow and design elements. In order to understand the conversational flow, some key terms need to be introduced:

Key elements to trigger interoperability between CA's and Chatbots

When designing queries for CAs there are several key terms and concepts to consider.

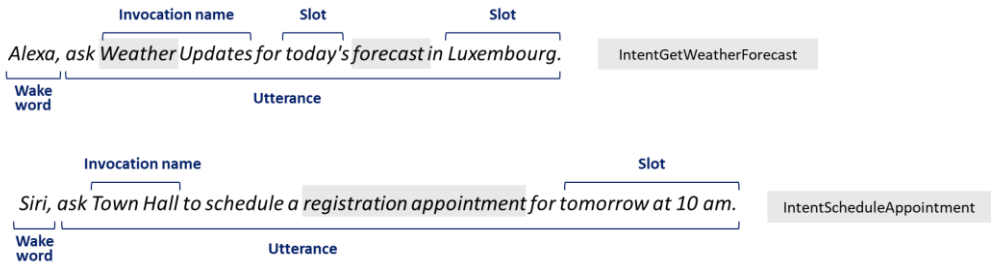
- Wake word: Specific word or phrase to call the CA into a listening mode ready to receive queries. The default wake word for CA's are usually the name, e.g., for Alexa devices it is "Alexa," but users can change it to other options such as "Echo," "Computer," or "Amazon." This word activates the device and lets it know that a voice command is coming. Some considerations to have with interoperability and wake words is the behaviour of using two wake word in one sentence (Mars, Feld, Rado, & Murphy, 2021).
  For instance, the following has two wake words: *"Alexa, what can Siri help me do?"* The desired output for this example would be for Alexa to be triggered and to reply. On the other hand, an undesired behaviour would be for Siri to be triggered.
- Invocation name: Specific word or phrase to activate a particular skill/action/shortcut. This can be used to trigger interoperability between CAs and Public Institution chatbots, by adding custom skills. Overall, an

[12] Change Siri accessibility settings on iPhone - Apple Support
[13] Type Requests to Alexa - Amazon Customer Service

invocation name like the name of the app within the CA environment. When a user wants to interact with a custom skill, they say the invocation name to open the skill. For instance, the invocation name 'recipe finder' is linked to a cookbook application, then the user can say 'Google assistant, ask Recipe Finder for a chicken recipe.'

- Intent: An intent represents the action that fulfils a user's spoken request. When users express different ways of asking for the same thing, those requests are mapped to a single intent. Intents help to define what the skill can do and what the user wants to achieve.
- Utterance: The spoken input from the user. These are the different ways users can phrase their requests to the skill. Utterances are defined within the skill's interaction model to map user requests to specific intents.
- Slot: Specific piece of data that can be extracted from the utterance to provide more information to fulfil the intent. Slots are used to capture important pieces of data that the skill needs in order to perform an action. Slots can be of various types, such as dates, times, numbers, names, etc.

Interoperability examples of Public institution skills added to CA's and that can be triggered through invocation names:



Transfer mechanism transparency

For redirection between CA's and relevant public Institution chatbot, several transfer mechanism options have been identified, see example in Figure 3. Transfer mechanism where assistant B would be the public institution chatbot.

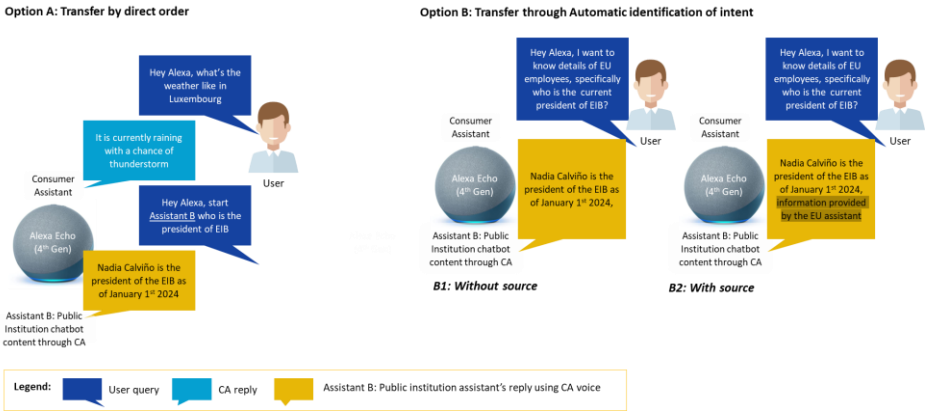


Figure 3. Transfer mechanism examples

Option A: Transfer by direct order

In this scenario, users can switch between CAs and public institution chatbots by using specific keywords. Keywords can be CA wake words, such as "Hey Alexa" and "Hey Google" or invocation words for a public institution chatbot with predefined skillset defined and downloaded in the CA, like "Assistant B". These activation phrases directly call the respective chatbot after being invoked through the CA, enabling it to respond to the user's query. This means that by using one CA device many different chatbots and the CA can be reached, the user just speaks to the CA and depending on the skill triggered, the right public institution chatbot will be called to provide the answer.

Option B: Transfer through Automatic identification of intent

In this scenario, users can switch between CAs and public institution Chatbots by using specific keywords that are linked to skills. Typically, these skills will be linked to invocation words for predefined skills, like, "Public services information" or "waste removal dates". The CA would then provide directly the answer by gaining the information from the source linked to the skill:

- Without source: Seamless interaction with answer directly provided by the CA.
- With source: Seamless interaction with answer directly provided by the CA including listing the source of the information (name of the CAs where the information was received).

Skill conversation lifecycle – context handling

The conversation between a CA and public institution chatbot through a skill can follow two different flows (described in step 3a-b)[14]:

1. A skill session begins when a user invokes the skill and Alexa sends the skills a request.
2. The skill receives the request and returns a response for Alexa to speak to the user.
3. Depending on the parameter selected in the skill response (during skill development) the next steps of the conversation could follow either approach:
    a. After Alexa speaks the content, the skill session ends. Alexa, not the skill, will therefore handle any further request unless the user re-invokes the skill (refer back to step 1 in this case).
    b. After Alexa speaks the content, the skill session stays open and the skill expects the user to respond. Alexa sends the new intent to the skill following process from step 2.

Example:

Option a: Simple flow - End of skill session

1. User: "Hey Alexa, start EU-who-is-who (invocation name) and tell me where the European Ombudsman is located?"
2. Skill through Alexa: "The European Ombudsman is located in Strasbourg, France."

Option b: Follow-up flow - Open skill session

1. User: "Hey Alexa, start EU-who-is-who (invocation name) and tell where the European Ombudsman is located?"
2. Skill through Alexa: "The European Ombudsman is located in Strasbourg, France."
3. User: "Give me the address"
4. Skill through Alexa: "1 Avenue du Président Robert Schuman"

[14] Manage Skill Sessions and Session Attributes | Alexa Skills Kit

It is important to note that in some complex cases, a CA may be unable to respond effectively to users. For example, Alexa might struggle to verbalize content if the information provided by a skill is too extensive to convey orally. Triggering document searches can be too complex for a voice assistant, which should be limited to Q&A.

User: "Hey Alexa, open Assistant B. Give me documents about climate change."

This request would likely be unmanageable for a device with only a speaker, as delivering multiple documents orally would be too difficult.

1.3.2.5 Consumer assistant search relevancy

When users seek information through consumer assistants like Alexa or Google Assistant, the results received are often curated from a variety of sources deemed reliable and authoritative.

Search engines such as Google or MS Bing retrieve their information through a complex process known as indexing, where they systematically crawl the web to gather and categorize content. Utilizing sitemaps can aid Google and other search engines in finding your content. This involves using web crawlers, or bots, that scan websites, and their data is stored in vast databases. Thus, one can break down the search process into three stages[15]:

1. **Crawling:** Automated programs called crawlers download text, images, videos, and other content from web pages.
2. **Indexing:** The content (text, images, videos) is analysed and stored in an index, which is a large database.
3. **Serving search results:** When a user searches, the search engine returns information relevant to their query.

For a public institution to ensure their website is selected as a source, they must focus on search engine optimization (SEO) practices to improve their site's visibility and accessibility. Some practices aim to address the challenge of improving website visibility in the absence of established methods to influence CAs toward reliable information retrieval. This includes some of the following best practices to improve a website visibility for the CAs:

Table 7. Improving websites' visibility

Best practice	Description to improve websites visibility for CAs
Eliminate hidden data	If a website contains hidden data, creating a separate version of the website without this hidden data can help make the information more visible and accessible. Overall, void having hidden data on your website to ensure all content is visible and accessible. This makes it easier for search engines to index the content for voice search queries.
HTML element	Making your website and links crawlable is to have a link with an anchor HTML element '<a>' with an 'href' attribute.
Use schema markup [16] Error! Bookmark not defined.[17]	Use structured data markup for rich results. Structure data is an essential SEO element as it makes it easier for search engines to interpret the search results accurately. Schema markup enhances the appearance of your page in search results examples can include article markup (detail information about an article with title, image, author, date) or organization markup that appear as panels on the search result page.

[15] In-Depth Guide to How Google Search Works | Google Search Central | Documentation | Google for Developers
[16] How To Optimize For Google Voice Search? 5 Best Tips!

Best practice	Description to improve websites visibility for CAs
Optimize for Mobile Error! Bookmark not defined.[16]	Ensure your website is optimized for mobile devices with fast loading times and easy navigation. Mobile-friendly websites are preferred by search engine crawlers.
Improve page speed[17]	Page speed is a critical ranking factor. Search engines prioritize websites that load quickly, which is especially important for voice search, where users expect instant answers.
Enhance Security (HTTPS)[17]	Upgrade your website to HTTPS to improve security. Websites that use HTTPS are favoured by search engines and users, leading to better rankings, including voice search results.
Create user-friendly site with quality content[18]	Creating high-quality, relevant content and using appropriate keywords to increase visibility, this can be done by: • Produce high-quality content favoured with descriptive and concise answers that can be easily delivered by voice**Error! Bookmark not defined.**[16] • Use long-tail keywords to capture specific voice search queries • Implement a FAQ section on your website to improve your chances of appearing in Featured Snippets, which are frequently used by Google Assistant for quick answers
Media support[18]	Leverage images and other media such as links to enhance your website's ranking potential. Well-optimized media helps improve user experience and can boost visibility in search results. Engaging in link-building strategies to gain backlinks from other reputable sites can also improve the institution's credibility and ranking.
Regular updates[18]	Keep your content regularly updated. Search engines favour frequently updated content, recognizing that you are actively maintaining and improving your site, which positively impacts rankings.
Backend agents	Utilizing metadata can increase visibility. Implementing an agent in the backend of the website to review and prioritize metadata can improve the website's crawlability, thereby making it more visible to CAs.
Crawl control	Manage and optimize how search engine bots navigate and index your website. This will ensure that search engines prioritize the most important content of a website while avoiding resource-intensive areas that may not need indexing (e.g., crawl rate, blocking parameters).
Pay	Consider paying an SEO agency (such as HigherVisibility, SmartSites)[19] for an in-depth analysis of your website. They can provide customized strategies to improve your website's performance and compatibility with voice search technologies (e.g., backlink analysis). Additionally, it is possible to pay for ads to appear on the first page of Google search. However, note that this method will not make the website more crawlable for a chatbot to use it as a search[20].
URL submission	Proactively providing search engines with the URLs of your website to ensure they are crawled and indexed. This helps search engines discover new or updated content quickly without waiting for their bots to crawl your site organically.

Formatted: Footnote Reference, Font: Not Bold

Formatted: Footnote Reference, Font: Not Bold

Formatted: Footnote Reference, Font: Not Bold

[17] Voice Search Optimization: 6 Tips to Improve Your Results
[18] How to Update Website Content for Better SEO
[19] Top 30 SEO Agencies - Jan 2025 Rankings | DesignRush
[20] Get your ads shown on the first page - Google Ads Help

Best practice	Description to improve websites visibility for CAs		
Specific to CAs	**Amazon Alexa:** Develop specific Alexa skills to enhance visibility and interaction opportunities. These skills will enable users to engage more effectively using voice commands as they are aware of the skill they trigger and the ultimate data source.	**Google Assistant:** Focus on optimizing content for Featured Snippets with clear and concise answers. This improves compatibility with Google Assistant, as these snippets are often used for voice responses.[21]	**Microsoft Copilot:** Offers Bing Webmaster Tools[22] to improve search relevancy (this package includes some of the above features, e.g., SEO reports, crawl control).

All these techniques can make website content more likely to be retrieved and utilized by search engines and CAs, such as Alexa and Siri. Currently, there are no definitive solutions to influence CAs to consistently retrieve information from reliable sources. Increasing your website's visibility for voice assistants involves a multi-faceted approach combining technical SEO practices and content optimization. By eliminating hidden data, using schema markup, optimizing for mobile, improving page speed, and enhancing security with HTTPS, you can create a strong foundation for better search engine visibility. Additionally, producing high-quality, concise content tailored for voice search, leveraging media support, and keeping your content regularly updated will further boost your site's appeal to voice assistants like Amazon Alexa, Google Assistant, and Microsoft Copilot. Finally, engaging a SEO agency for a detailed analysis can provide specialized strategies tailored to maximize your website's performance in the evolving landscape of voice search technology.

This section explored UX considerations when interacting with CAs and strategies for public institutions to enhance their visibility in CA search results. The next section will delve into viable approaches to facilitate interoperability with CAs.

1.4 Viable approaches

1.4.1 Feasible options to facilitate interoperability

This section explores viable methodologies to establish interoperability of a public institution chatbot with a CA. First, the study will investigate the different ways in which the chatbots can interoperate looking at different approaches. Then a comparison will be established for the different approaches to compare.

1.4.1.1 Approach A – Skill based transfer

It is generally easier to implement one-way/ unidirectional interoperability flow where the assistant can be accessed via the CA. In this approach, the CA acts as the intermediary where the Public institution chatbot skill is called. The reply is provided back through the CA (usually responding in the CA's voice. Even though customisation to other voices can be made according to **Error! Reference source not found.**~~1.3.2.3~~ **Error! Reference source not found.**~~Communication aspects~~).

Basic conversational flow

[21] How to Optimize for Voice Assistants (Alexa, Google Assistant, Siri) - Bird
[22] Bing Webmaster Tools

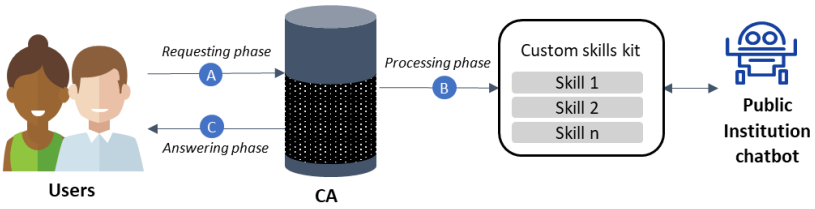


Figure 4. Skill based transfer

A. Requesting phase: User sends requests to your assistant through the CA (skills set up through API or SDK - additional information on these capabilities in 1.3.1.1).

B. Processing phase: The CA would be triggered based on the invocation name of the skill (see 1.3.2.4 Conversational flow elements) and forward the request to the public institution's assistant for a response.

C. Answering phase: This response is retrieved and sent back through the CA using the CA capabilities (e.g., Alexa/ Siri's voice).

Set up and requirements to facilitate *Skill based transfer* interoperability

This section delves deeper into *Approach A Skill based transfer integration* by outlining the necessary preconditions for integration. It will also provide a detailed examination of each phase presented and a simplified overview of the functional architecture.

Technical set-up requirements

- **CA device** (this can be any CA speaker or other application that can be the entry interface for the user).
- The vendor should offer a **CA Kit** or similar functionality to enable custom developments of skills. These skills will often be a voice-driven capabilities enhancing the CA functionality with skills from public institution chatbots or other. This can be cone through SDK or API provided by major CA platforms (see 1.3.1.1 Technology considerations~~Technology considerations~~).
- **Preconditional steps**
  - Assistant API: A functional assistant with an accessible API endpoint to handle requests
  - Skill Development and Deployment: The skill must be developed, registered, and deployed on the corresponding CA platform. Skill should be assigned to a specific invocation name tailored for the public institution chatbot
  - Skill accessibility: Ensure the skill is published and accessible. Users must download the skill, provided it is available in their region (to be specified when publishing the skill).

The diagram below shows the detailed flow of the processing phase to illustrate the use of the different components of *Skill based transfer* interoperability. Examples are used in the description below to make the processing clearer.

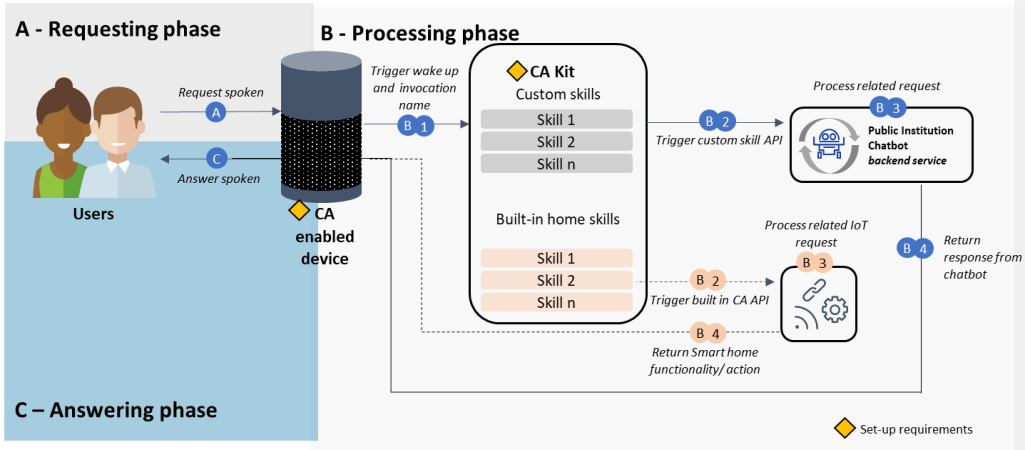


Figure 5. Overview of Skill based transfer architecture

The use of an CA functionality for interoperability, shown in Figure 5~~Figure 5~~~~Error! Reference source not found.~~**Error! Reference source not found.**, works as follows:

A. Requesting phase: The user speaks their query to the CA, starting with the *Wake Word* and asking the question using an *Invocation Name* of the skill

> *e.g., a user asks about public service hours in their region to Siri: 'Siri, which public services are open in Luxembourg city on the weekend and what are their opening hours?'*

B. Processing phase:

1) The CA recognizes the wake word and subsequently the invocation name for the skill (in our case it would be a custom skill that would be facilitate interaction with another chatbot in this example interoperability with a public institution chatbot)
2) The CA device recognises this as a custom skill and triggers the related API connecting to the public service chatbot.
3) The user's request is sent to the backend endpoint
    i. The skill's backend processes the incoming request.
    ii. Extracts the intent and parameters from the request.
4) The public institution chatbot's API processes the request, generates a response, and sends it back to the CA.

> *e.g., invocation name 'public service' is triggered with the specific ask for 'weekend operating times in the Luxembourg city'. The CA device (Siri in this example) recognises this as a custom skill, triggers the API connecting to the public service chatbot, which in turn understand the intent and returns the response containing the operating hours to Siri via API.*

C. Answering phase: The CA that was the entry point for the user to ask the question, responds to the user.

> e.g., Siri responds with *'The following public services are open Saturdays from 9:00 AM - 5:00 PM in Luxembourg city: Public Library, Community Center, Local Health Clinic & Post office. All of them are closed Sundays.'*

1.4.1.2 Approach B – Gateway based transfer

Gateway based transfer offers a seamless UX between assistants. This is a setup where interaction passes through the CA device application (e.g., Alexa Echo or Siri) and either the public institution chatbot or the CA can reply with its own capabilities offering a multi agent experience. This is driven by a middleware (referred to as an **interoperability library** in this study) that exposes different chatbot APIs to enable control of multiple VAs though one interface. Some vendors[23] have a portable setup that easily enables this transfer, but it can also be developed from scratch.

In case both Siri and Alexa as well as a Public Institution Chatbot is integrated in the CA interoperability Library, it would enable users to gain information form any of the above parties through the entry device.

Basic conversational flow

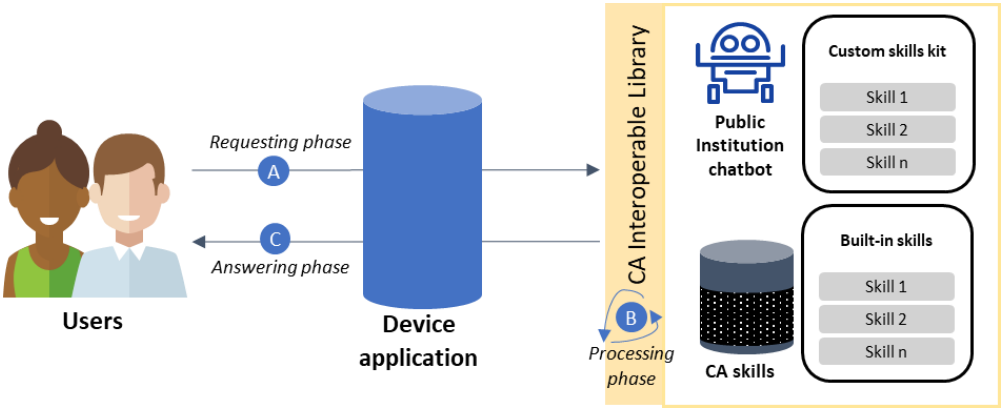


Figure 6. Gateway based transfer

A. **Requesting phase:** User requests would be sent to a device application.

B. **Processing phase**: The interoperability library would be triggered based on the wake word and after gaining the relevant intent forward the request through the library's dialogue/ action manager to the relevant assistant for a response. The response is received in the transfer manager component of the interoperability library.

C. **Answering phase:** This response is sent back through the device application using the capabilities of the triggered assistant.

Set up and requirements to facilitate *Gateway based transfer*

This section delves deeper into *Approach B - Gateway based transfer* by outlining the necessary preconditions for integration. It will also provide a detailed examination of each phase presented and a simplified overview of the functional architecture.

Technical set-up requirements

- **CA device**: Refers to any device that can be used to call the assistants, for instance this can be an Echo device from Alexa, or a phone using Siri, a website or application with a CA.

[23] AutoGen Studio: Interactively Explore Multi-Agent Workflows | AutoGen, max-toolkit/README.md at main · alexa/max-toolkit · GitHub

- **Interoperable library:** An interoperable library is a collection of resources to manage multiple assistants and handle their interaction patterns. Depending on the CA selected the interoperable library could be available pre-built[24] (see 1.3.1.1 Technology considerations~~Technology considerations~~). The library is typically divided into two parts:
    - **Application side requests –** Ensures the interoperable library is triggered and the correct functions are activated
        - **Wake word observer:** Detects the specific word or phrase that wakes up the voice assistant.
        - **Experience Manager:** Manages the overall UX, ensuring smooth interactions (e.g., ensures actions that overlap are triggered in the right order).
        - **Physical listening controller:** Controls the hardware related to listening, such as microphones.
    - **Assistant side –** Links to all assistants in the interoperable network
        - **Dialogue/ action manager:** Oversees the conversational flow and actions based on user input by triggering the correct intents from the relevant assistant backend.
        - **Transfer manager:** Manages the transfer of tasks or data between different VAs and shares the final answer/ action to the device application.
- **Preconditional steps**
    - Defined Interoperability Framework: Existence of a standardized framework and API provided by the interoperable Library. Set up pre-build interoperable library (if available through vendor of CA in use) or build this component.
    - Roles and responsibilities**:** Define the roles and responsibilities of creating and maintaining the interoperable library.
        - **CA:** If the interoperable library is pre-built and provided by the CA, they would be responsible for maintaining the parts related to the CA updates.
        - **Public institution:**
            - If the interoperable library is pre-built by the CA, the public institution would be responsible for maintaining aspects related to their chatbots.
            - If the public institution builds the interoperable library, they are solely responsible actor for maintaining it.
    - Set up and expertise: Integrate all relevant assistants into the interoperable library to enable the interoperable experience. Enable APIs that can coordinate the *Gateway based transfer* experiences both visually and audibly through Universal Device Commands (UDCs) (Lantin, 2021). This would require:
        - Hardware Compatibility: Ensuring that the hardware can support the integrated setup.
        - Software Ecosystem: Availability of supported software and development environments for integrating multiple assistants.
        - Developer Expertise: Access to skilled developers experienced in multi-agent systems and integration methodologies.
        - Agent: Already have an agent in development or production.

It is important to note that the Chatbots & CA's never directly call each other, they always make use of the *universal interoperability library* for use cases that require interaction. The diagram below shows the detailed flow of the processing phase to illustrate the use of the different components of *Gateway based transfer*. Examples are used in the description below to make the processing phase clearer.

[24] GitHub - alexa/max-toolkit: The MAX Toolkit provides software which aims to accelerate the development of devices which integrate multiple voice agents. The Toolkit provides guidance to both device makers and agent developers towards this goal.

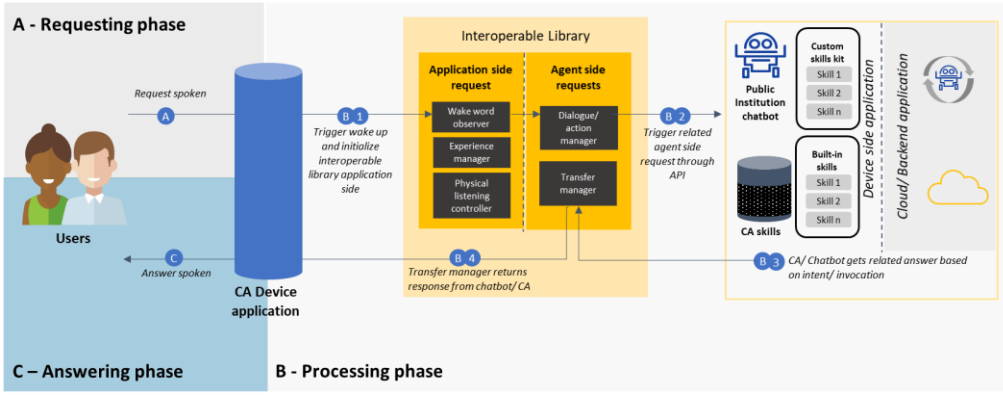


Figure 7. Overview of Gateway based transfer architecture

The use of a CA functionality for interoperability, shown in **Error! Reference source not found.Error! Reference source not found.**~~Figure~~ 7, works as follows:

A. Requesting phase: The user speaks a query to the CA device, starting with wake word/ the invocation name of the assistant it wants to be directed to (other options possible, see transfer mechanism in 1.3.2.4)

e.g., a user says, 'Hi Alexa, when is the next date for recycling trash removal in my street?'

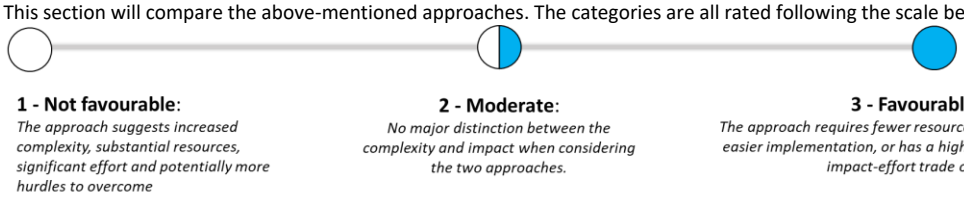
B. Processing phase:

1) The CA Device received the query and wake up word triggers the Interoperable Library (application side) to listen.
   e.g., In the above example 'Alexa'
2) The Interoperable Library (application side) receives the user query and determines which Chatbot or CA needs to be involved. This triggers the Dialogue/ action manager from the Interoperable Library (agent side) to call a specific relevant agent based on the intent/ feature of the query.
   e.g., In the example above, the public Institution chatbot covering the intents for 'recycling trash removal'.
3) The interoperable Library coordinates the interaction between the Device application and the called chatbot.
   i. The Device application uses the Interoperable Library's client-side APIs to send the query to the called assistant.
   ii. The called assistant processes the request received from the Interoperable Library (forward the request to its cloud-based backend service, if necessary, to fetch the required information).
   iii. The generated response is sent to the Interoperability Library (Agent side) transfer manager component.
4) The Interoperability Library (Agent side) transfer manager component sends the response to the CA Device application.

C. Answering phase: The Device Application speaks the reply (by using text to speech) to the user.
e.g., Alexa responds with *'The next recycling trash date is tomorrow. Please ensure to have all glass and paper waste in the allocated location before 08:00am.'*

1.4.2 Comparison of the different approaches

This section will compare the above-mentioned approaches. The categories are all rated following the scale below:



1 - Not favourable:
The approach suggests increased complexity, substantial resources, significant effort and potentially more hurdles to overcome

2 - Moderate:
No major distinction between the complexity and impact when considering the two approaches.

3 - Favourable:
The approach requires fewer resources, easier implementation, or has a higher impact-effort trade off.

Figure 8. Comparison scale

When comparing the interoperability between CAs and public institution chatbots, it is important to consider the various evaluation criteria.

Table 8. Comparison of the different approaches

	Approach A: Skill based transfer		**Approach B: Gateway based transfer**	
C1 - Set-up complexity	●	Simple set up, requires definition of skill and validation with CA skills system.	○	Requires set up and integration with an Interoperable Library (configure each assistant, ensure communication protocols are established, follow toolkit framework provided for software and hardware, etc.)
C2 - Maintenance	◑	Low maintenance once skill is set up. Only adaptions required if skill and/or the assistant's API changes.	◑	Low maintenance once set up. If changes in the skill, adaptions required on functioning of the interoperable library.
C3 - User awareness	○	User needs to use specific invocation name to trigger skill and download before the skill to be able to have a direct interaction via the CA.	●	Seamlessly integrated and managed by the interoperable library allowing interactions with multiple assistants through a single device.
C4 - Scalability	◑	While it can handle moderate scaling, introducing new assistants would require implementing a new skill for each assistant.	◑	Can integrate multiple assistants in one device, making it adaptable to future expansions and additional functionalities as needed. However, each assistant needs to be included in the interoperable library.
C5 - Response time	●	Generally, provides fast responses as it is a direct communication between the CA and the skill's backend.	○	Potential delays due to the additional layer in overall flow of communication as the Interoperable Library might introduce some latency.

	Approach A: Skill based transfer		Approach B: Gateway based transfer	
C6 - Transparency	◐	Users have downloaded the skills and are aware of how the skill should be used, as they should learn the invocation name and coverage as part of the skill installation process	◑	Depends on the transfer mechanism (see options for transfer transparency in 1.3.2.4 Conversational flow elements).

By evaluating these factors, stakeholders can make informed decisions to achieve optimal interoperability depending on their current setup and needs. Overall Approach A is simpler and easier to integrate, while approach B would create a more seamless UX when set up. Both approaches have benefits and drawbacks, and the decision will depend on the organisations needs and technological setup as well as the maturity of the chatbot and the skill that is required to be integrated with a CA. The next section will focus on having a regulatory outlook when setting up interoperability with CAs.

1.5 Regulatory outlook

1.5.1 A view on relevant EU regulations

In the context of interoperability between assistants, several key regulations must be considered, many of which have already been addressed in the *Interoperability between public institution chatbots publication* (chapter added in Appendix). When we consider interoperability, the different organisations involved in the interoperability network would need to be considered as separate data controllers and as such transparent to the end user on how data is processed or shared is paramount. All parties in the interoperable network have to accept and consider the reputational risk to work with other parties in case of any data breach or non-compliance. When integrating public sector chatbots with CAs, specific regulatory aspects of specifically the GDPR[25] can be considered.

General Data Protection Regulation (GDPR)

Use of personal data and data sharing

In the context of integrating a public institution chatbot with a CA like Alexa, Siri or Google assistant, handling personal data becomes especially relevant due to the personalized nature of services provided by CAs. Since CAs often collect, process, and utilize personal data to offer tailored experiences, the interoperability of these systems necessitates stringent controls and clear boundaries on data sharing.

Not all personal data should be shared automatically between the two systems; instead, data sharing should be governed by principles of *data minimization* and *purpose limitation* to comply with the EU's GDPR. This means that only the personal data necessary to fulfil the specific user request should be shared, and users must be fully informed and *provide explicit consent* for their data to be used across both platforms. Additionally, implementing robust data protection strategies, such as pseudonymization and encryption during data transfer, is crucial to safeguarding user privacy and ensuring that personal data is not exposed to unauthorized access or misuse. By establishing these boundaries, both systems can offer a seamless integrated experience while respecting user privacy and adhering to EU regulatory requirements.

[25] Regulation - 2016/679 - EN - gdpr - EUR-Lex

As we have seen with the two approaches in section 01.4 where personal data is managed differently. Furthermore, additional considerations need to be taken into account as with interoperability conversations and data exchanged through the CA device may be stored and processed.

Specific considerations regarding the Always-Listening Features in Consumer Assistants

CAs are triggered with Wake words, such as "Morning Alexa" for Amazon devices, or "Hi Siri" for Apple devices. Legally, these devices are required to respect certain privacy, data protection standards and the legal requirements. Wake words have the following operational implications:

- **Background Listening:** While the device listens constantly in the background for the wake word, it typically should not store or process any conversation information until the wake word is detected. This mode of operation respects the principle of data minimization.
- **Wake Word Activation:** Once the wake word is detected, the device actively processes the following requests. The processing of data begins only after the wake word is identified, mitigating unnecessary data collection. Transparency should be included through comprehensible privacy notices that should explain to the users that their device listens for specific wake words and what happens when the wake word is detected.
- **User Control:** Users typically have control over their data through device settings. This control allows them to delete recordings or turn off the always-listening feature if and when they choose to. Information on how to disable the always-listening feature or customize settings should also be easily accessible.

EU AI Act

General overview

The interoperability between CAs like Alexa, Siri, Google Assistant, and public institution chatbots would be impacted by the EU AI Act[26]. The Act aims to ensure the safe and ethical deployment of AI technologies across the EU by imposing strict regulations on transparency, accountability, and risk management. CAs and public institution chatbots will need to comply with these regulations, which could affect how they handle data, maintain privacy, and ensure security during interactions. For instance, the Act's requirements on high-risk AI systems necessitate thorough assessments to mitigate potential risks and bias. An AI system is considered high-risk if it is used in a specific category such as biometric identification and categorization of natural person, education and vocational training, etc. Consequently, organizations must implement rigorous compliance measures, such as conducting impact assessments, ensuring transparency in data processing, and establishing clear accountability frameworks. Most often, the EU AI Act might impose transparency obligations on chatbots to ensure that end-users are aware they are interacting with an AI system. These regulatory demands may lead to increased operational oversight and necessitate technological adjustments to align with the EU's stringent AI governance standards. Overall, adhering to the EU AI Act will help foster trust and reliability in AI-driven interactions between CA and public institution chatbots.

Interoperability specifications and compliance

Since CA would be considered the providers in scenarios where public institution chatbots reply through CA devices, the compliance landscape under the EU AI Act is influenced accordingly. If the public institution ensures compliance with the AI Act and their chatbot skill is validated during the CA's review process, it indicates adherence to both regulatory standards and the CA's policies. Given that the CA's policies are designed to comply with the EU AI Act, this dual-layered compliance implies that the public institution's chatbot services provided through the CA would

[26] Regulation - EU - 2024/1689 - EN - EUR-Lex

also be compliant. Therefore, by meeting the rigorous evaluation criteria of the CA and adhering to the stipulated regulatory requirements, public institutions can ensure that their AI-driven interactions delivered via CAs remain aligned with the EU AI Act, thereby maintaining legal and ethical integrity.

Over and above these measures, companies may need to conduct Data Protection Impact Assessments (DPIAs) to evaluate and mitigate privacy risks associated with always-listening features. This would assess the necessity and proportionality of the data processing activities, especially continuous waiting for the wake word. Adequate security measures must be implemented to protect the data processed after detecting the wake word. This includes protection against unauthorized access and ensuring data integrity.

1.6 Implementation framework

This chapter outlines the structure for creating an end-to-end interoperability PoC, linking a public institution chatbot to a CA. Following the proposed phases and using these templates will support interoperability PoC development according to current best practices and guidelines from earlier phases of the study. Appendix B1~~B1~~ contains a table of anticipated deliverables for an interoperability project, each with scope and content examples that can evolve. The framework guides public institutions in connecting their assistants to CAs, overcoming challenges, and moving towards a more interoperable network. Each phase describes the goal, key prerequisites, deliverables, and adjustments based on the selected approach from 1.4.1. General guidelines, templates, and checklists are provided to support the PoC from design through testing to implementation.

Formatted: Not Highlight

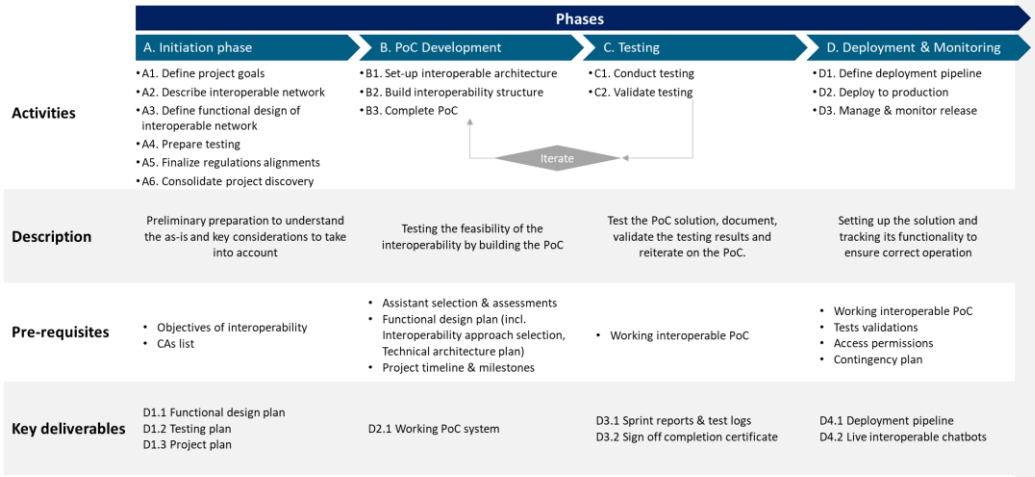


Figure 9. Four phases to implement an interoperability PoC between public institutions chatbots with CAs

1.6.1 Phase A: Initiation

Goal: The goal of the initiation phase is to define the scope of interoperability with CAs. This phase includes all aspects of project discovery to determine which CA to consider, how the integration will be facilitated, which technologies will be used, etc.

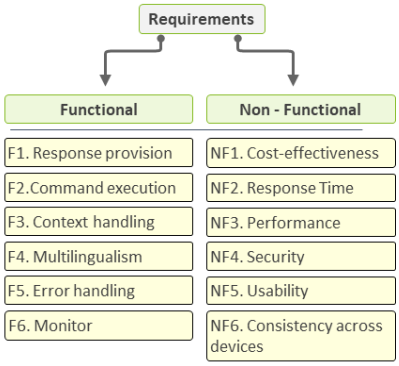
Activities: The following key activities are proposed to cover all aspects of the project initiation phase.

Table 9. Phase A: Activities descriptions

Activity + Description	Deliverables	Related Sections
A1. Define project goals Consider the following aspects: • **Select relevant CA to connect to:** Conduct extensive research on the available CAs, their capabilities, and their compatibility possibilities (Approach A vs. Approach B). • **Define user base:** Comprehend citizens and requirements and expectations from the assistants to ensure that the services provided meet user needs. • **List augmentation benefits**: List how the CA interoperability can enhance or add to your services.	Key features list (part of D1.1: Functional design plan)	1.2.2~~1.2.1~~, 1.2.3~~1.2.2~~, 1.4.1, 1.4.2
A2. Describe interoperable network • **Define the conversational flow:** Define the conversational flow and transfer mechanism approach – *the selection of the transfer mechanism might influence the decision of the interoperability approach to select and/or vice versa*. • **Identify the conversational requirements**: Define the conversation scenarios and the dialogues required to handle them (e.g., invocation names and intent types as well as the actions linked to these). • **Define target users**: Define which region of the world would be relevant to make the interoperability accessible to and the marketing strategy to share the interoperability.	Key features list (part of D1.1: Functional design plan)	1.3.2.4
A3. Define functional design of interoperable network • **Define functional and non-functional requirements:** Functional requirements of an interoperability system guide its operational features, while non-functional requirements relate to system quality, ensuring efficiency, maintainability, and reliability. • **Select an interoperability approach:** The selection of the approach will define the mechanism for including public institution chatbots into CA's. To select can be done depending on need of project (e.g., a more seamless interoperability) and resources available. ○ Approach A – Skill based transfer: In a skill-based transfer the public institution must create a skill to connect through the CA selected. Creation of a skill might slightly differ from CA to CA as seen in 1.3.1.1. ○ Approach B – Gateway based transfer: In this approach the connection between the assistants is done through an interoperable Library. • **Select transfer mechanisms:** Select your transfer mechanisms options for assistant interoperability *(e.g., see* Figure 3*)* • **Define architecture and technology stack**: Identify the required technology components to support interoperability keeping the requirements set-up in the previous activity in mind – *the components required will differ based on the interoperability approach selected (e.g., interoperability library or CA Kit)*.	D1.1: Functional design plan *(incl. technical architecture)*	1.3.1, 1.3.1.1, 1.3.2, 1.3.2.4, 1.4.1 Appendix B.2.1

Activity + Description	Deliverables	Related Sections
A4. Prepare testing • **Prioritize requirements:** Determine requirements' significance and their testing sequence, usually based on business value, risk, complexity, and impact. • **Define epics:** Define large units of work called Epics which can be disaggregated into smaller tasks based on the prioritized pre-defined requirements. • **Define User Personas:** Define fictional characters representing actual users and their behaviour, used for guiding design or test decisions. • **Draft User Stories:** Narratives illustrating users' perspective of interacting with the product, created based on the user personas. **Validate User Stories:** User stories are approved and signed off by stakeholders.	D1.2: Testing plan (incl. prioritization, epics, user stories, etc.)	Appendix B.2.2
A5. Finalize regulations alignments **Comply to regulations:** Legal compliance - Make sure you are compliant with existing laws, such as the AI Act and GDPR (and other regulations if you are outside of the EU)	n/a	1.5.1
A6. Consolidate project discovery • **Draft the project plan:** Timeline for the PoC development, the milestones, foreseen meetings and stakeholders involved. **Set acceptance criteria and Definition of Done (DoD):** The acceptance criteria is used to determine if the PoC meets the desired goals and requirements. DoD defines criteria to meet for the PoC to be considered complete *(e.g., assistants to provide answers in both English and a French, UAT completion).*	D1.3: Project plan Acceptance criteria and DoD (part of D1.2: Testing plan)	

Templates to accelerate phase A



When investigating interoperability with CAs, it is crucial to understand and identify the functional and non-functional requirements that will serve as foundations in the development, deployment and maintenance of the system. Some of these will feed to UX considerations (topics covered in sections 1.3.2). Below are some examples of functional and non-functional requirements. Additional requirements should be added and adapted to project goals[27] (more detailed requirements' description available in Appendix B.2.1~~B.2.1~~)

Formatted: Not Highlight

Figure 10. Example of functional & non-functional requirements

Regarding the testing preparation, User Acceptance Testing (UAT), also known as UAT or End-User Testing, is the last phase of the software testing process. During UAT, the software is tested by the real users who will be using the software in the real-world environment. UAT is important because it helps in validating that the system is ready

[27] Not all will be relevant, and the interoperability case might require additional ones

for release. It confirms that the system meets the agreed business requirements and can handle tasks in real-world scenarios according to the specifications. The main purpose of UAT is to ensure that the software system is working as expected before it's moved into the live environment. If any issues or improvement areas are identified, it gives the development team an opportunity to resolve them based on the user's feedback. Figure 11 shows an overview of the UAT testing which will be explained in more details in the appendix B2.

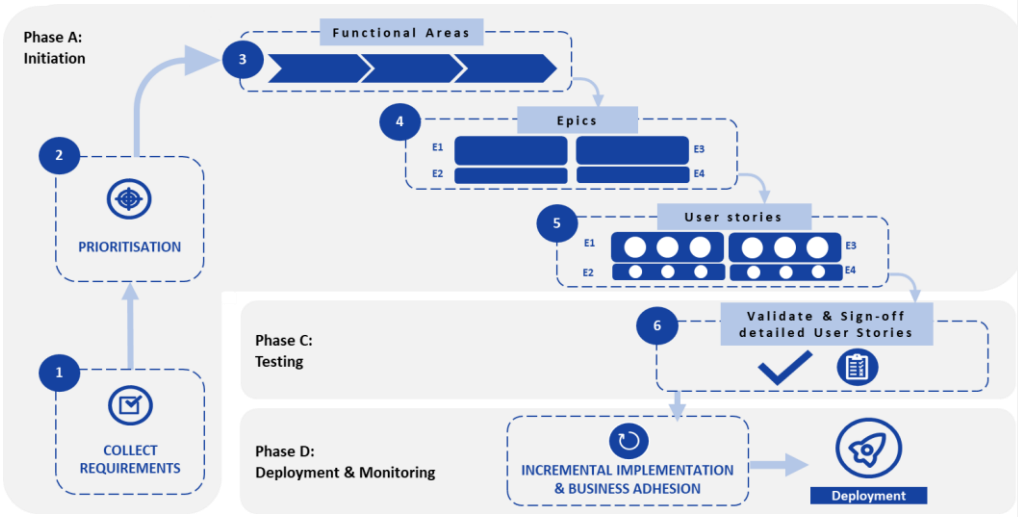


Figure 11. Overview of UAT testing process

A detailed description of steps 4, 5, 6 (Epics, User Stories, Validate & Sign-off detailed User Stories) is available in the Appendix B2.

1.6.2 Phase B: PoC Development

The PoC development aims to showcase functionality and assess the practicability of the interoperability with CAs before building a complete system. It helps anticipate potential issues and establish expectations for the final product. The upcoming section will detail an agile implementation for the PoC development, covering setup of interoperable architecture, conversation element design, infrastructure building, and documentation. The PoC development will be an iterative process adapted based on testing phase outcomes.

Table 10. Phase B: Activities description

Activity + Description	Deliverables	Related Sections
B1. Prerequisites / Set up interoperable architecture • **List and get all necessary components, tools and accesses**: Assemble and secure all necessary components (e.g., hardware compatibility, software ecosystem), tools (e.g., API) and access permissions for stakeholders – this might include the creation of a developer account from the CA selected depending on the approach picked.	-	1.3.1.1 1.4.1

Activity + Description	Deliverables	Related Sections

B2. Build interoperability structure
• **Create a skill or set up an interoperability library:** Depending on the approach selected either the creation of a custom skill or building/ using a pre-built interoperable, if provided by the CA, library will be necessary.
• **Develop specific trigger mechanisms:** Based on the design and defined conversational flows, develop the trigger mechanism and conditions based on user interactions.
  ○ Set an invocation name or wake word.
  ○ Define the trigger mechanism, see Figure 3.
• **Backend Endpoint**: Connect your assistant by adding your API endpoint.
	-	1.3.1.1
		1.3.2.4
		1.4.1

B3. Complete PoC • **Refine PoC:** Iteratively optimize and refine the PoC with the test phases (see phase C) and fix discrepancies identified in testing. • **Validate PoC:** Make sure acceptance criteria and DoD have been reached (e.g., defined response time). Hereafter, testing takes place and the full Skill/ feature is deployed as per Phase D.	D2.1 Working PoC system	Phase C 1.6.3

This section ventured through the PoC development, reviewing the key stages necessary to bring the concept into reality. These insights derived will be instrumental in guiding the direction of subsequent Testing and Development sprints. Concluding with Phase B, the PoC Development, and transitioning into Phase C, Testing. These two phases move in lockstep, each influencing and shaping the other in a continuous feedback loop (sprint). The results from the Testing Phase directly feed into the next sprint of the PoC Development, allowing for precise and focused refinements. Conversely, the newly advanced output of the PoC Development is subjected to evaluation in the Testing Phase. It is through this iterative process that we take steady strides towards the realization of the PoC.

1.6.3 Phase C: Testing

This section deals with the evaluation of the interoperability with CAs. The goal of this phase is to evaluate the functionality, reliability and efficiency of the interoperability. The table below goes into further details for each phase. For templates on how to accelerate phase C, consult Appendix B3~~B2~~.

Formatted: Not Highlight

Table 11. Phase C: Activities descriptions

Activity + Description	Deliverables	Related Sections
C1. Conduct testing • **Perform tests:** Test the predefined test cases prepared in A4 to ensure the PoC does not have defects or unexpected behaviours. The test cases can be categorized as "met/not met/potential to meet". Depending on the approach selected the testing could be done on the CA testing console or simulator, see 1.3.1.1. • **Log defects:** Document the defects seen in the testing, it is important to keep track and reproduce the defect to fix it. • **Prioritize**: Based on the priority matrix of the requirements, the defects will be assigned a scope and score to define the priority of the defect (low to high). They can also be prioritised for sprint inclusion (e.g., Defect ID 2 – low priority - will be looked at in sprint 2) • **Fix:** Based on the reproduced errors and priority level, find and solve the defects. Confirm fix by reproducing the same defect with the same conditions. • **Scope of next sprints & retest:** The testing and PoC development are iterative and will take place in several sprints. The defects fixed as well as other requirements to build will be part of the next sprint to develop in the PoC and test again.	D3.1 Sprint reports & test logs	Section: 1.3.1.1 Phase A 1.3.1.1~~1. 3.1.1~~ Phase A
C2. Validate testing • **Deliver and Iterate UAT:** The UAT process is reiterated until all requirements are covered and no more issues (moderate and high) are discovered with the PoC before the final delivery and the DoD has been reached. • **Validate:** Validate testing by relevant stakeholders.	D3.2 Sign off completion certificate	Section: Phase B

Concluding the testing phase, this approach ensures that the interoperable network meets the business needs and works as expected in real-world scenarios. It is iterative, responsive, and user-focused making it a key element in successful PoC development. After the successful completion of testing, the next phase focuses on integration into the user environment with continuous tracking.

1.6.4 Phase D: Deployment & Monitoring

The goal of the deployment and monitoring phase is to ensure the smooth integration and interaction of the assistants within the user environment, while continuously supervising the system to confirm that the solutions is working well. This phase also aims to use the monitoring insights to optimize and improve the interaction between the assistants based on user feedback and system observations.

Table 12. Phase D: Activities descriptions

Activity + Description	Deliverables	Related Sections
D1. Define deployment pipeline (might differ depending on Approach) • **Define necessary deployment components:** Key components are necessary to set up the deployment pipeline such as Version Control System (VCS), build server, automated testing, artifact repository, deployment automation. • **Commit stage:** Trigger commit code to the VCS. This will allow code changes to be fetched from VCS and for the build server to compile and run the code. • **Automate testing:** Execute various test to ensure code functionality, and reliability (e.g., unit tests, integration tests). • **Set-up and deploy to staging testing:** Once all testing have been validated, create a staging environment that will look like the production environment. Deploy the solution there, this will allow to test under real-world conditions providing a final check before the production. • **Beta version:** Depending on the approach and CA selected, you could have the option to deploy the interoperability feature in a beta version where it can be distributed to a limited number of users. • **Implement stability or stress testing:** Before launching, the interoperable system must be thoroughly stability tested for handling varied real-world scenarios, peak loads, high data volume, and sustainable performance, preventing crashes, slow responses or data loss when used by consumers. • **Validate with stakeholders:** Get final approval from stakeholders to move to the production deployment.	D4.1 Deployment pipeline	-
D2. Deploy to production (might differ depending on Approach) • **Define production deployment strategy (if necessary):** Define and select deployment strategy such as canary releases, blue-green deployments. • **Deploy to production environment:** Trigger deployment to production environment. • **CA review:** Depending on the approach once you deploy your skill in production through the CA console platform, a review process will be initialised. • **Deployment status:** Once reviewed and approved by the CA company, your bot will be available for users through the CA in the region you have selected. • **Versioning:** When updating information or making some changes in a new version of your project, the new version will have to be resubmitted and go through a full review process before being released again to the public.	D4.2 Live interoperability assistant	-

Activity + Description	Deliverables	Related Sections
D3. Manage & monitor release • **Monitoring tool:** Depending on the approach and CA selected a monitoring tool might be available to monitor analytics on the interoperability. • **Define Key Performance Indicators (KPIs):** List measurable KPIs that tie in with business objectives and customer expectations. Relevant KPIs could be response time, user satisfaction rate, accuracy of information, number of successful interactions, or problem resolution rate. • **Provide continual assessment:** KPIs should be perpetually tracked and evaluated from the moment of deployment and throughout the system's lifecycle. This enables prompt detection of any potential issues or discrepancies and allows for timely rectification. • **Implement performance analysis:** Collect and analyse data on individual assistant performance within the interoperable network. This helps to identify any weak links and optimize overall system performance. • **Optimize:** Apply insights derived from monitoring and analysis to continuously refine and optimize the interoperable assistants, align closely with KPIs, and improve UX.	-	

Monitoring and KPIs play an integral role in maintaining the efficacy and efficiency of the interoperable assistant system. They not only indicate the current performance level but also provide valuable insights highlighting areas for improvement. Regular monitoring backed by robust KPIs ensures constant system optimization, improved user satisfaction, and effective error handling. In a rapidly evolving AI landscape, it is this iterative cycle of monitoring and evaluation that ensures the assistants remains effective, accurate, and user centric.

1.6.5 PoC Implementation framework

1.6.5.1 PoC Scope and Description

The PoC was launched with the objective of testing the interoperability between public institutions' chatbots and consumer assistants. For this purpose, a multilingual voice assistant application connected to the Publications Office chatbot was developed for Amazon's Alexa customer assistant using the Alexa Skills Kit (ASK).

The skill provides users with access to key information covering EU Law in Force, EU Publications, EU Whoiswho, published by the Publications Office of the EU through text and natural voice interactions in English and French, across all supported locales. The system maintains in-session conversation history to enable follow-up questions without needing users to repeat prior context, mimicking behaviour found in standard chatbot applications. The implementation is unidirectional where the Q&A API from the Publications Office sends information to the custom Alexa Skill.

Several user experience (UX) features have been implemented to make interactions smoother and more natural:
- Progressive responses to keep users engaged during long API calls.
- Fallback intent handling and helper messages to guide users through possible interactions.
- Custom cards sent to the Alexa app for visual reinforcement of voice responses.

From a monitoring perspective, CloudWatch logs and Amazon Developer Console analytics were used to track skill performance, latency, and user behaviour, providing valuable insights for iterative improvement.

The PoC was tested across multiple platforms to ensure compatibility and broad accessibility, including Alexa-enabled devices (an Echo 1st Generation was used to test device compatibility) and, for mobile users, the Alexa app.

Seven acceptance Criteria have been defined to evaluate the PoC, based on functional and non-functional criteria. The table below details the functional and non-functional requirements, outlining the technical considerations and coverage within the PoC. The evaluation of the acceptance criteria against three success levels is presented as part of the PoC results.

Table 13. Functional and Non-Functional Requirements for the PoC

	ID	Criteria	Description	Technical consideration	Technical coverage
Functional requirements	**F1**	**Response provision**	Both Alexa and Alexa using OP portal knowledge through Publio Q&A API, should be able to answer user queries through Alexa device with Alexa skills.	Out-of-scope queries of Publio Q&A API will not be answered by Alexa.	Out-of-scope questions response mechanism.
	F2	**Command execution**	Alexa skills should be able to understand user intent and provide Publio API Q&A answer to user. No blocking point in flow should occur where user is stuck (Alexa should guide user if no direct answer is obtained).	Error-handling flow to be implemented to avoid Alexa unresponsiveness.	A fallback intent was integrated to handle user requests not understood by Alexa.
	F3	**Context handling**	When Alexa skill uses OP portal knowledge through Publio Q&A API, context refers to the conversation history that is understood by Publio Q&A API for multiple follow-up questions. Alexa skills and Publio/OP portal usage reacts consistently with regards to context.	API context handling is limited to 3 previous user-chatbot interactions. Alexa will use the Publio Q&A API response provision, and no additional contextualization will be implemented to ensure consistency with Publio and Portal.	The Alexa skill can handle context up to three previous interactions by using the built-in API history cache feature.
	F4	**Multilingualism**	Alexa skill is developed in French and English – given device setup for language.	• Language cannot be switched in flow, either one language or other used per conversational flow (linked to the device set language)	Selected user stories were tested in each selected language, with one tested in both for

				• PoC relies on Alexa built-in speech-to-text understanding for words in foreign languages (and people speaking in different accents).	consistency checks.
Non-functional requirements	**NF1**	**Response time**	Alexa should be able to respond to user queries within acceptable time limits (Alexa skill relies on API response provision and Alexa should not add additional latency).	• Limited to Publio Q&A API response time that sometimes exceeds 10 seconds.	Progressive responses were implemented to keep the user engaged during the waiting time. They inform the user that the connection with OP was successful, and include an EU fun fact.
	NF2	**Performance**	Corresponds to how well the Alexa and the Publio Q&A API interact and if the correct chatbot responds to the user query (looking at the effectiveness and quality of the output).	Monitoring of PoC performance relies on statistics tracked in AWS services. Access to the general skill analytics are open for OP. AWS logs only accessible by Deloitte – shared upon request.	• 'Analytics' tab in Developer Console (contains metrics on skill and intent activations). • AWS Lambda logs for each user interaction.
	NF3	**Usability**	**Intuitive usage:** Using Alexa skills to access OP portal knowledge through Publio Q&A API should be user-friendly and intuitive for users. **Action execution:** Skills are multi-device (i.e., can be used both via Echo and App). If using Alexa App, actions are also available and functioning.	Once the OP skill invocation name is triggered, prior to a query an intent activation utterance or full sentence is required (e.g., "*Tell me […]", "What […]"*).	• User manual delivered with testing phase. • Multiple intent activation phrases defined during development to ensure natural and user-friendly intent recognition.

1.6.5.2 Timeline

The table below presents an overview of the project timeline, including an outline of the activities performed during each phase of the PoC and its deliverables.

Table 14. Overview of PoC project timeline.

Phase & key activities	Deliverables	Timeline
Project Management	-	April – June 2025 (3 months)
PHASE A: Discovery • Setup project: PoC Scope • Conduct PoC workshops ✓ *WS1 - Scope confirmation UX* ✓ *WS2 - Technical architecture* ✓ *WS3 - Testing setup*	DLV1: UX alignment deck	April 2025 (1 month)
	DLV2: Technical architecture deck	
PHASE B: Delivery • Environment setup • Develop PoC • Testing (two sprints) • Defect fixing	DLV3: Functioning system	April 2025 – June 2025 (2 months)
PHASE C: Closure & Documentation • Document technical specifications • Validation & handoff	DLV4a: Technical documentation	May 2025 – June 2025 (2 months)
	DLV4b: Testing user manual	
	DLV5: Demo presentation	
	DLV6: Final report	
	DLV7: Project closure presentation	

1.6.5.3 Testing

UAT session 1: overall outcome

For the first UAT, 6 testers (4 for English, 2 for French) were present and all six User Stories were tested according to the predefined languages and test cases, each testing a feature or functionality of the developed Alexa skill. As a result, 10 tickets were opened based on their feedback relating to defects and changes identified during this session.

Main achievements:

- Improvement of UX by improving the Q&A intent handling in the skill (i.e., since we were in dedicted skill we were able to resource to a simple “Tell me…” or “What is…” activation which made the conversation with the customer assistant feel more natural)
- Identification of limitations related to the Alexa system and services (e.g., speech-to-text accuracy for foreign names and acronyms, sensibility to accents and environment noise)

- Improvement of the *fuzzy matching* technique for increasing the accuracy of the name recognition (sample of 50 names of the EU Whoiswho database was used)
- Identification of API-related limitations (e.g., lack of response for queries that should be in scope for the Q&A API[28])

UAT session 2: overall outcome

For the second UAT, 4 testers were present (who already participated in UAT1 and already had an initial understanding of the PoC as well as a comparison from both UAT) and tested the same User Stories as in UAT 1. In total, 4 new tickets were opened based on their feedback relating to defects and changes identified during this session.

Main achievements:

- Confirmation that the enhancements developed after the feedback in UAT 1 were successfully implemented (e.g., improved *fuzzy matching*, new Q&A handling feature)
- Improvement of Q&A capability by expanding the activation phrases in French, to account for the language's diverse expressions and variations

Overall, more tests were performed by more users in TS1 and the following week – This is due to a higher volume of defects and additional developments performed after the feedback from the first testing session.

1.6.5.4 Results

The below table shows the outcome through the Acceptance criteria and status outcome from UAT sessions. In green the Met acceptance criteria and in Yellow the Potential to meet Acceptance criteria.

Table 15. Evaluation of Acceptance Criteria (AC) for the PoC

Legend:

Met acceptance criteria
Potentially to meet acceptance criteria – to be further refined in industrialization

[28] These have been logged for a subsequent industrialization phase of the Q&A API, as they are out of scope for the Task 02 PoC development.

ID	Name	Description	Not met (1)	Potential to meet (2)	Met (3)	Linked Req.
AC1	**Skill recognition**	Both Alexa and Alexa using OP Portal knowledge through Publio Q&A API, should be able to answer user queries through Alexa device with Alexa skills.	Alexa never recognizes the OP Alexa skill invocation name (neither in EN or FR).	Alexa only sometimes recognizes the OP Alexa skill invocation name.	Alexa always recognizes the OP Alexa skill invocation name (in EN & FR).	[FR.01]
AC2	**Command execution**	Alexa skills should be able to understand user intent and provide Publio API Q&A answer to user. No blocking point in flow should occur where user is stuck (Alexa should guide user if no direct answer is obtained).	Alexa does not understand the intent or does not redirect to OP Alexa skill.	OP Alexa skill sometimes understands the intent and provides the correct answer (e.g., gives the phone number).	OP Alexa skill understands the intent and provides the correct answer.	[FR.02]
AC3	**Context handling**	When Alexa skill uses OP Portal knowledge through Publio Q&A API, context refers to the conversation history that is understood by Publio for multiple follow-up questions. Alexa skills and Publio/OP portal usage reacts consistently with regards to context.	In the OP Alexa Skill, no context is considered; the Alexa skill directly switches back to Alexa or does not consider any previous question through the Q&A API Cache history feature.	In the OP Alexa Skill, the context is sometimes considered OR context is only considered for two or less previous prompts. Context is not consistent to Publio context handling.	In the OP Alexa Skill, the context is considered for the three previous prompts (Alexa uses the Q&A API response provision, to ensure consistency with Publio).	[FR.03]
AC4	**Multilingualism**	Alexa skill is developed in French (US5) and English All USs) – given device setup for language.	The OP Alexa Skill does not answer queries in the language queries when	The OP Alexa Skill answers some queries in the language queries where asked when the Alexa devices is set to	The OP Alexa Skill answers all queries in the language queried when the Alexa devices is set to that	[FR.04]

			the Alexa devices is set to that language.	that language. OR this functions only in either English of French.	language (for both EN and FR).	
AC5	**Response time**	Alexa should be able to respond to user queries within acceptable time limits (Alexa skill relies on API response provision and Alexa should not add additional latency).	When calling the OP Alexa skill, the Alexa skill adds additional latency to the API Q&A response and has no progressive response to decrease user's experience of the latency.	When calling the OP Alexa skill, there is sometimes progressive responses to decrease user's experience of the latency (but not always when the response is slow).	When calling the OP Alexa skill, no additional latency is added to the Publio Q&A API response and progressive responses is always present to decrease the user's experience of the latency.	[NFR.01]
AC6	**Performance**	Corresponds to how well the Alexa and Publio interact and if the correct chatbot responds to the user query (looking at the effectiveness and quality of the output).	No analytics is available to track and monitor the OP Alexa skill performance.	Analytics is available to track and monitor some interactions of the OP Alexa skill performance.	Analytics is always tracked to monitor interactions of the OP Alexa skill performance. After the OP Alexa skill is triggered, intents are recognized and Q&A outputs shared by Alexa through the Publio Q&A API for relevant prompts.	[NFR.02]
AC7	**Usability**	**Intuitive usage**: Using Alexa skills to access OP Portal knowledge through Publio Q&A API should be user-friendly and intuitive for users .	No user guide is provided, and users have difficulty to access and set up OP Alexa skills.	User guide is provided, but not clear and detailed, some users have difficulty to access and set up OP Alexa skills.	User guide is provided and complete. Users easily follow instructions to access and set up OP Alexa skills on their devices.	[NFR.03]

Overall, the PoC showed satisfactory results for the majority of defined Acceptance Criteria, satisfying both functional and non-functional requirements set up at the start of the project. Below are the most relevant findings from the testing sprints and key limitations and improvements identified.

Intent activation and usage

The statistics on intent activation during the testing sprints show that following the introduction of the new Q&A handling feature after the first testing session, the OP skill's user experience had a substantial enhancement:

- Reduction of fallback (i.e., unrecognized intents) from 20% to 12%
- Increase of Q&A intent usage and recognition (from 56% to 75%)

Figure 12. Intent recognition and usage by testing sprint, average for all languages

Nonetheless, there are differences in intent recognition accuracy between languages. During the second sprint, fallback was higher in French (16% vs 9% in English), reflecting the language-specific difficulties of the Alexa voice transcription service.

Figure 1313. Intent recognition and usage by language, Sprint 2 (04/06-11/06)

Skill and API latency

During both testing sprints, the latency of the skill and the underlying API were recorded in order to estimate the waiting times experienced by users at each Q&A interaction with the skill.

From the analysis of the interactions performed during the week of the testing sessions, the results indicate that the average waiting time for one question asked to the skill is around 9-10 seconds. This time includes on average

6 seconds of waiting time due to API latency (i.e., the time the API takes to return an answer), and the remaining 3-4 seconds for skill backend processing (e.g., speech-to-text processing, intent recognition, eventual fuzzy matching). However, the API latency is quite variable, as in some instances responses are very quick (i.e., less than a second), while other times they can take up to 30 seconds.

These findings indicate that the introduction of the progressive responses to mitigate the skill latency are indeed an helpful tool to reduce the impact on user experience.

Table 16. Skill latency breakdown by testing sprint

	Sprint 1 (19/05-26/05)		Sprint 2 (04/06-11/06)	
	Average	**Range**	**Average**	**Range**
API time	5,0 sec	0,8 - 17,1 sec	7,2 sec	0,8 - 27,7 sec
Skill time	3,8 sec	3,5 - 4,1 sec	3,4 sec	3,2 - 4,0 sec
Full Q&A time	8,8 sec	4,6 - 20,9 sec	10,6 sec	4,1 - 31,0 sec

1.6.5.5 Limitations and mitigations

Throughout the development and testing phases of the PoC, several limitations were identified. The following outlines the key technical and platform-related challenges encountered, along with the mitigation strategies that were implemented to minimise their impact on the user experience.

API-Related Limitations

- **Latency:** The Q&A API occasionally exhibits high response times, with replies typically averaging around 10 seconds, but ranging from as low as 3 to over 30 seconds in some cases. To mitigate the user impact of this delay, progressive responses were implemented. These help keep users informed and engaged while the final response is being prepared.
- **Unresolved requests due to API timeouts:** As the Q&A API response time is quite variable, the Lambda function timeout (i.e., the maximum time for which a skill request will hang before returning a timeout error that closes the skill session) is set to 40 sec, meaning that for a request that is still processing over this time, the skill will return an error and close the session since it received no response from the API. At timeout, Alexa will say "*The skill has encountered an issue"* and the skill session will be closed. This is the automatic timeout behaviour from Amazon and cannot be changed. To avoid this scenario, a second timeout parameter has been set for the API request itself. This has been set to 35 sec and, if reached, the skill would return an error message and prompt the user to ask again. These settings ensure that the API is not frequently timed out and the user experience is not severely impacted by the high waiting times.
- **Answer Accuracy:** In accordance with the project's assumptions, the answers provided by the Q&A API are used as-is, with no modification or post-processing on the skill side. However, source formatting is improved within the skill to present references in a clearer, more user-friendly manner.

Amazon Alexa Platform Limitations

- **Speech-to-Text Accuracy:** Alexa's voice recognition struggles with certain foreign names, acronyms, and abbreviations – for example, "*Jan Plánovský*" may be misinterpreted as "*Joan Planos*". To address this, *fuzzy matching* was introduced using a sample from the EU Whoiswho people list, increasing the chance of correctly identifying people names despite transcription errors. Acronyms, on the other hand, were excluded from the processing with the same technique as their introduction increased the chance of incorrect matches (e.g., the query "*Who are the assistants of Hilda Herman*" became "*Who are the Eurostat of Hilde Hardeman*" after *fuzzy*

matching on people names and acronyms). As there are workarounds for increasing the transcription accuracy of acronyms (e.g., spelling the name), they have been excluded from the processing.

- **Constraints for Invocation Name and Intent Activation Utterances:** Alexa's naming policy imposes restrictions on invocation names in certain languages. Additionally, pronunciation ambiguities in French, particularly with plural words, can make it difficult for Alexa to recognize the skill name accurately, resulting in activation issues. To address this, the French skill invocation name had to be carefully adapted to meet policy requirements while remaining recognizable. Furthermore, one activation phrase for the Q&A intent could not be used in French due to contractions and hyphenation ("*Y a-t-i-l {query}*", i.e. "*Is there {query}*"), forcing the use of alternate question triggers (e.g., "*Existe-t-il {query}*") for Q&A activation. Any future adaptations or industrial-scale deployments should account for these constraints.
- **Lack of Mid-Conversation Language Switching:** Alexa does not support switching languages dynamically during a session. The skill will operate only in the language configured in the user's device/app settings. This limitation cannot currently be mitigated, as it stems from constraints within the Amazon Alexa platform.
- **Feature Deprecation and Privacy Restrictions:** Amazon's increasing focus on privacy has led to the removal or restriction of certain functionalities (e.g., card notifications no longer trigger active phone alerts). The skill was designed within these constraints to still offer a meaningful user experience, without compromising privacy principles.
- **Intent Activation Usability:** Activating the Q&A functionality requires the user to say a specific activation phrase (e.g., *"What…", "Who…", "Where…",* etc.). However, if the user utterance does not contain any of the predefined ones or if Alexa incorrectly transcribes the speech, the intent will not be recognised and the skill will return a fallback response. Mitigation efforts include:
    - Adding fallback guidance that prompts users to say the correct phrase (i.e., instructing the user to start the question with *"Tell me…"*).
    - Designing short, memorable utterances and incorporating natural entry points like "*Tell me…*", "*I have a question…",* or "*Do you know…",* on top of the existing natural question starters (*"What…"*, etc.).
- **Fixed Response Style:** The skill is configured to use a simple response style to suit voice-based interaction. While users may ask for more advanced explanations (e.g., "*Can you explain in expert terms?*"), the API response style cannot be dynamically changed at runtime. This behaviour is fixed during the skill build and cannot currently be overridden.

1.6.5.6 Next steps for industrialization

Following the successful implementation and evaluation of the PoC, several next steps have been identified to guide the path toward production readiness and industrialization. These are grouped into key thematic areas, ordered by priority:

- **Managing API Latency:** While progressive responses offer a viable workaround for end users during the PoC, this solution is not scalable for industrial-grade applications. A long-term improvement plan will be required, potentially involving optimizations on the Q&A API side or exploring alternative architectures to reduce response time.
- **Expanding Language Support:** To better serve a broader audience, future iterations of the skill could support additional languages beyond English and French, depending on Alexa's locale availability and project scope. Note that the addition of languages does not involve changes on the skill's source code, but only inclusion of new configurations (i.e., Alexa response messages and intent activation phrases) for the target language.
- **Addressing Speech Recognition Issues for Foreign Names:** A persistent challenge has been Alexa's limited ability to accurately transcribe foreign names. The following actions are being considered:

    - **Monitoring Amazon Updates:** Stay informed of potential improvements to the Alexa Voice Service that may address current transcription shortcomings.
    - **Enhanced Name Matching in Industrialization:** For full-scale deployment, the integration with the EU Whoiswho database could be expanded beyond the reduced dataset used in the PoC. This will include:
        - **Improved fuzzy matching algorithms** to better interpret incorrectly transcribed names.
        - **Investigating additional techniques** (e.g., phonetic matching or AI-based correction) to further boost recognition accuracy.
- **Skill Certification:** The immediate next step involves preparing the skill for submission to the Alexa Skill Certification process. This includes final testing, metadata preparation, privacy compliance validation, and aligning with Amazon's publishing requirements across supported locales.
- **Feature Enhancements:** Further feature development could be considered, such as memory expansion (e.g., storing more context beyond the current three-turn history) or enabling personalized interactions and preferences, in alignment with evolving user needs and privacy policies.
- **Cross-Platform Testing and Support:** Exploring the potential for testing and deploying the Q&A experience on other virtual assistant platforms (e.g., Google Assistant) could offer additional flexibility and broader accessibility.
- **Security Considerations:** Ensuring the continued protection of user data and compliance with EU regulations will remain a top priority. Future iterations of the skill may explore:
    - **Data Privacy by Design**: Continue incorporating privacy principles into the updated architecture, ensuring minimal data collection and proper anonymization where possible and needed.
    - **Session Management**: Enforcing strict session validation, token expiration, and user authorization checks to prevent session hijacking or unauthorized access to the skill or API.
    - **Regular Security Audits**: Conducting periodic penetration tests, vulnerability scans, and code reviews to proactively identify and address security risks in the skill's backend.
    - **Compliance Monitoring**: Ensuring alignment with EU digital policies, including data localization requirements, accessibility standards, and cybersecurity best practices.

1.6.5.7 Cost estimates for an industrialization phase

Following the successful completion of the Proof-of-Concept (PoC), cost projections for the industrialization phase of the project have been simulated based on actual AWS billing data and usage patterns observed during the PoC.

During the three-month period of activity, the costs associated with the Publications Office skill were categorized into fixed and variable.

Fixed costs, averaging about $53.50 per month, covered essential baseline services such as encryption (e.g., AWS Key Management Service) and security monitoring (e.g., AWS Security Hub).

Variable costs depended on how frequently the skill was used and included activity-based charges for cloud function execution (e.g., AWS Lambda, EC2), log and data storage (e.g., Amazon S3), system monitoring (e.g., Amazon CloudWatch), and network access (e.g., AWS Virtual Private Cloud). For the PoC phase, there was an average of $39.60 per month.

It is important to note that in AWS's pay-as-you-go model, the distinction between fixed and variable costs is not always strict. In this context, the fixed costs refer to recurring baseline expenses that don't fluctuate significantly

with minor usage changes, but still increases as the infrastructure expands. On the other hand, variable costs scale more directly with usage volume (e.g., the number of skill invocations or the duration of compute activity).

As the system transitions into the industrialization phase, with usage increasing from ~1,000 monthly invocations during the PoC to 100,000 invocations per month, the cost structure is expected to evolve as follows:

- Fixed costs are suggested to conservatively budget at $510/month. This increase is to provision the larger infrastructure for all base components for a more robust and scalable production setup. The growth includes expanded logging and storage capacity (e.g., via Amazon S3, ~$15/month), more extensive monitoring and alerting configurations (AWS CloudWatch and AWS CloudTrail, ~$50/month), and enhanced encryption and compliance operations (AWS Key Management Service, ~$150/month; AWS Security Hub, ~$75/month; AWS GuardDuty, ~$15/month; AWS Config, ~$5/month), and storage or other components needed on production scale. These infrastructure-level upgrades are essential for supporting a high-availability, production-grade Alexa skill.
- Variable costs are estimated to stabilize at around $430/month, even as usage grows 100-fold. This is due to efficiencies at scale such as lower per-request costs for cloud function execution, which significantly reduce the average cost per invocation. While the PoC incurred $0.03–$0.05 per request, the industrial phase is expected to average ~$0.004329 per request, representing a nearly 9-fold reduction in cost per invocation.

The total monthly cost during the industrialization phase is therefore estimated at ~$940, combining both fixed and variable components. While this marks an overall increase in monthly expenditure compared to the PoC, the sharp drop in per-request cost highlights the system's scalability and cost-efficiency at higher volumes.

These estimates are intentionally conservative to account for potential infrastructure overhead and service expansion, offering a realistic and robust financial outlook for scaling the skill into production.

Table 17. Overview of PoC AWS infrastructure costs and forecast for an industrialization phase.

				Forecast (for industrialization)					
	Apr'25	May'25	Jun'25[30]	Month1	Month2	Month3	Month4	Month5	Month6
Fixed Account Cost, ($)	53.5	53.5	53.5	510	510	510	510	510	510
Variable Cost, ($)	37	52.2	29.7	430	430	430	430	430	430
Total cost ($)	90,5	105,7	83,2	940	940	940	940	940	940

[29] The variable cost per invocation is calculated as the total variable cost per month divided by the total monthly invocations count ($430/100.000). Variable costs include services such as VPC provision (required for the API connectivity) and Lambda invocation price, including related EC2 instance usage. To be conservative, one of the largest Amazon EC2 instances (*t3a.xlarge*) was selected for the industrialization cost forecast.
[30] Values as per June 25th, 2025.

1.7 Conclusion

The interoperability between public institution chatbots and widely available CAs holds the promise of delivering trustworthy information to a broader audience, thereby enhancing public service delivery and increasing user adoption.

Key findings

Our research indicates a rising trend in initiatives centered around multi-agent conversations, especially those incorporating LLMs. While integration with CAs is feasible, it faces limitations due to ongoing major developments in CAs aimed at incorporating LLMs, often leading to the discontinuation of previous functionalities. Additionally, there are currently no established interoperability standards.

Several key considerations are highlighted and thoroughly detailed in the study. From a technology standpoint, the considerations include evaluating currently available technologies versus those being phased out. Monitoring options are another critical aspect, dependent on whether they are already available on the CA platform. Contractual considerations and certifications necessary for publishing on the CA are also discussed.

UX considerations includes the importance of user awareness and visibility, ease of use, and localization options, which can be selected for some CAs. Communication aspects for interoperability with CAs must account for the channels—voice and text—depending on the CA device users are utilizing. This also involves addressing available languages, covering both the languages supported by the CA and the public institution. Additionally, the tone of the CA is crucial, involving decisions on whether the public institution's voice can be used (if already implemented in the chatbot), the CA's voice, or a combination of both. Other findings from the study show that crucial elements in CAs and interoperability concern conversational elements such as wake words, invocation names, or intents which trigger the CA and would also help in triggering the interoperability. Finally, an important aspect of this study was looking at how public institutions can make their information more visible to CAs.

The study also focused on viable approaches to achieve interoperability. Two primary approaches were identified. Using the CA SDK, this approach involves incorporating the public institution chatbot as a "skill" of the CA, which can be called based on an invocation name. The second approach uses an interoperable library that allows multiple agents to be integrated. The library could be provided by the CA, depending on the specific CA in question. The preferred approach can be selected based on various factors including the CA to be integrated, setup complexity, scalability, transparency, and more.

Finally, the study addressed the regulatory aspects, focusing on two key regulations: the EU AI Act and GDPR. The main takeaways include data sharing considerations between the CA and the public institution chatbot, and the necessity of maintaining transparency throughout the process. Our research underscores the potential and challenges of achieving effective interoperability between public institution chatbots and consumer assistants, paving the way for more inclusive and efficient public service delivery.

Future directions

Our study underscores both the promise and challenges of achieving effective interoperability between public institution chatbots and consumer assistants (CAs), paving the way for more inclusive and efficient public service delivery. As the domain of multi-agent conversations evolves—especially with the growing role of generative AI—several key directions emerge for future development and industrialization.

- **Generative AI Integration**
  Exchanges with CA providers indicate that generative AI is the next significant step for enhancing consumer assistants. Some platforms are already adopting these technologies, which hold potential for enabling more human-like conversations, better understanding of user intent, and improved multilingual capabilities.

- **Improving Interoperability and Standardization**
  While efforts toward interoperability between CAs, multi-agents, and institutional systems have existed, they remain limited in scope. The study reveals several viable approaches, with the choice depending on factors such as the number of chatbots to connect, integration methods supported by CAs, setup complexity, and user awareness. There is a clear need for more standardized protocols and frameworks to streamline these connections across different platforms.
- **Multilingual Support**
  Supporting multiple languages is central to making these systems more inclusive. Expanding beyond English and French will be important in future iterations, depending on locale availability and technical feasibility within each CA platform.
- **Name Recognition and Speech Accuracy**
  A recurring challenge lies in the accurate transcription of foreign names by voice assistants. Improving this will require:
    - Staying informed about advancements from CA platforms that may enhance recognition.
    - Incorporating more robust name-matching strategies, including phonetic or fuzzy matching.
    - Exploring integration with larger or more authoritative databases, like the EU Whoiswho, to better interpret names during interactions.
- **Security and Privacy**
  As these systems scale, ensuring user data protection will remain a top priority. Future developments should continue to incorporate privacy-by-design principles, secure session management, and compliance with EU regulations on data handling, accessibility, and cybersecurity.
- **Cross-Platform Potential**
  While the focus has been on Alexa in the current phase, extending the solution to other virtual assistants (such as Google Assistant) is worth exploring. This would allow for broader accessibility and test the adaptability of the interoperability framework.

In conclusion, the evolution of interoperability between public institutions and consumer assistants is well underway, with generative AI, improved standards, and multilingual support setting the stage for more seamless and citizen-centric services. This study not only highlights current gaps but also offers a forward-looking perspective to guide the next phase of development.

2 Appendix

A. References

Alepis, E., & Patsakis, C. (2017). Monkey says, monkey does: security and privacy on voice assistants. *IEEE Access, 5*, 17841-17851.

Alexa unveils new speech recognition, text-to-speech technologies. (2023, September 23). Retrieved from Amazon science: https://www.amazon.science/blog/alexa-unveils-new-speech-recognition-text-to-speech-technologies

Apple. (2024, June 10). *Introducing Apple Intelligence, the personal intelligence system that puts powerful generative models at the core of iPhone, iPad, and Mac*. Retrieved from Apple: https://www.apple.com/newsroom/2024/06/introducing-apple-intelligence-for-iphone-ipad-and-mac/

Budzinski, O., Noskova, V., & Zhang, X. (2019). The brave new world of digital personal assistants: Benefits and challenges from an economic perspective. *NETNOMICS: Economic Research and Electronic Networking, 20*(2), 177-194.

Connect a bot to Alexa. (2022, December 09). Retrieved from Microsoft: https://learn.microsoft.com/en-us/azure/bot-service/bot-service-channel-connect-alexa?view=azure-bot-service-4.0

Conversational Actions. (2024, September 18). Retrieved from Google Assistant: https://developers.google.com/assistant/conversational/build

End of support for Cortana. (2024). Retrieved from Microsoft: https://support.microsoft.com/en-us/topic/end-of-support-for-cortana-d025b39f-ee5b-4836-a954-0ab646ee1efa

European Parliament & Council. (2009). Directive 2009/24/EC of the European Parliament and of the Council on the legal protection of computer programs (Codified version) (Text with EEA relevance). *Official Journal of the European Union*.

Glover, E. (2024, August 20). *24 Popular AI Assistants*. Retrieved from builtin: https://builtin.com/artificial-intelligence/ai-assistant

Glover, E., & Whitfield, B. (2024, August 06). *23 Popular AI Assistants*. Retrieved from Builtin: https://builtin.com/artificial-intelligence/ai-assistant#:~:text=AI%20assistants%20often%20utilize%20natural,responses%20in%20a%20conversational%20manner.

Hoy, M. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly, 37*(1), 81-88.

Kharpal, A. (2024, July 10). *Samsung to launch upgraded voice assistant Bixby this year with its own AI*. Retrieved from CNBC: https://www.cnbc.com/2024/07/11/samsung-to-launch-upgraded-bixby-this-year-with-its-own-ai.html

Klein, A., Hinderks, A., Rauschenberger, M., & Thomaschewski, J. (2020). Exploring Voice Assistant Risks and Potential with Technology-based Users. In *WEBIST* (pp. 147-154).

Koetsier, J. (2021, June 28). *Alexa, Siri, Google Assistant: How The Top Smart Assistants Stack Up*. Retrieved from Forbes: https://www.forbes.com/sites/johnkoetsier/2020/08/08/alexa-siri-google-assistant-how-the-top-smart-assistants-stack-up/

Mars, R., Feld, E., Rado, R., & Murphy, J. (2021). *Voice Interoperability Initiative Architecture Series Whitepapers: Voice Input Data Handling on Multi-Agent Devices.* Amazon, Inc. or its affiliates. Retrieved from https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers-Voice_Input_Data_Handling_on_Multi-Agent_Devices.pdf

McLean, G., Osei-Frimpong, K., & Barhorst, J. (2021). Alexa, do voice assistants influence consumer brand engagement?-Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research, 124*, 312-328.

MobiDev. (2022, March 21). *AI Virtual Assistant Technology Guide 2022*. Retrieved from Medium: https://mobidev-biz.medium.com/ai-virtual-assistant-technology-guide-2022-c375d7539aa0

Overview. (n.d.). Retrieved from Bixby Developer Center: https://bixbydevelopers.com/dev/docs/get-started/overview

Oyston, G. (n.d.). *Chatbot vs. AI Agent: What's the difference and why does it matter?* Retrieved from ada: https://www.ada.cx/blog/chatbot-vs-ai-agent-what-s-the-difference-and-why-does-it-matter/#:~:text=Chatbots%20follow%20scripted%20conversation%20workflows,respond%20and%20action%20customer%20queries.

Pessarlay, W. (2024, July 19). *Samsung brings native LLM into play to advance voice assistant Bixby*. Retrieved from Coingeek: https://coingeek.com/samsung-brings-native-llm-into-play-to-advance-voice-assistant-bixby/

Rausch, D. (2023, September 20). *Previewing the future of Alexa*. Retrieved from Amazon: https://www.aboutamazon.com/news/devices/amazon-alexa-generative-ai

Thormundsson, B. (2023, December 05). *Most used voice assistants in the United States in 2022, by device*. Retrieved from Statista: https://www.statista.com/statistics/1368148/voice-assistants-use-by-device-united-states/#:~:text=In%202022%2C%20there%20was%20a%20diverse%20use%20of,a%20healthy%20spread%20among%20all%20day-to-day%20use%20devices.

What is the Alexa Skills Kit? (2024, July 01). Retrieved from Amazon.com: https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html

What is the Alexa Skills Kit? (2024, July 01). Retrieved from developer amazon: https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html

Wiggers, K. (2024, february 2024). *Google Assistant is now powered by Gemini — sort of*. Retrieved from TechCrunch: https://techcrunch.com/2024/02/08/google-assistant-is-now-powered-by-gemini-sort-of/

Wood, C. (2024, April 27). *10 Top AI Assistants*. Retrieved from VKTR: https://www.vktr.com/ai-platforms/10-top-ai-assistants/

Yaghoubzadeh, R., Kramer, M., Pitsch, K., & Kopp, S. (2013). Virtual agents as daily assistants for elderly or cognitively impaired people: Studies on acceptance and interaction feasibility. In *Intelligent Virtual Agents: 13th International Conference, IVA 2013, Edinburgh, UK, August 29-31, 2013. Proceedings 13* (pp. 79-91). Springer.

Zhou, T. (2023, September 21). *How do voice assistants like Alexa & Siri actually work?* Retrieved from Medium: https://medium.com/@zhouxiaogan0/how-do-voice-assistants-like-alexa-siri-actually-work-1885dce1f683

B. Additional Content

B1. Overview of the potential deliverables of the implementation framework

The table below presents the anticipated deliverables to be generated over an interoperability between Public institution chatbots and CAs. Each deliverable outlines the scope and content it should include, these remain examples and can be enhanced as the project evolves.

Table 18. Detailed overview of the potential deliverables

Phase	ID	Deliverable	Description (purpose)
Phase A	D1.1	Functional design plan	Plan to outline the functional infrastructure for deploying the assistants, this includes the necessary functional design for interoperability, conversational flow & trigger mechanism (CA selection, directional flow, requirements, intent, etc.). This also include key features list comparing different possible features to include in the interoperability, covering: • List CA considered (with their capabilities, topics, etc.), the user based targeted, the potential benefits for both assistants • Assistant and interoperability features to consider between the assistants, the intents, inappropriate topics, conversational flow, list of languages of the assistants • This should be done considering B.2.2~~D.3.2~~ user stories • The technical architecture plan with the technical infrastructure to set-up for the PoC development (data structure, translation layer, routing interactions, host solutions, configuration, architecture diagram, etc.)
	D1.2	Testing plan	Document describing: • The requirements (functional and non-functional) prioritization by importance of tasks and issues to oversee in the interoperability between the assistants. • The epics and related user stories linked to the epics, including the personas • Acceptance criteria & DoD: list the acceptance criteria (e.g., response speed, conversational robustness) and the DoD to be reached to consider the PoC complete and move to the production deployment.
	D1.3	Project plan	Document detailing the project work plan with the timeline for the PoC development, the key milestones, the risks and mitigations, foreseen meetings and stakeholders involved.

Phase B	D2.1	Working PoC system	The complete working system should reach the DoD defined and the acceptance criteria set.
Phase C	D3.1	Sprints reports & test logs	Reports documenting the tests achieved, the defects identified, their prioritization and fixes applied as well as the scope of the next sprint.
	D3.2	Sign off completion certificate	Signed off completion certificate by stakeholders providing validation of UAT (all test validated in regard to the acceptance criteria and DoD). This certificate is necessary to move to the production deployment.
Phase D	D4.1	Deployment pipeline	Document detailing the steps of the deployment pipeline from the necessary components, environment, additional testing to the final approval.
	D4.2	Live skill / interoperability library	Fully functional interoperable deployed assistant in production that can interoperate on the CA devices.

B2. Implementation framework: Templates for Phase A – Initiation

B.2.1 Functional / non-functional requirements description

When investigating interoperability between public institution chatbots and CAs, it is crucial to understand and identify the functional and non-functional requirements that will serve as foundations in the development, deployment and maintenance of the system. Many of these will feed to UX considerations or the approaches to select to achieve interoperability which are topics covered in sections 1.3.2 & 1.4.1. Below are some examples of functional and non-functional requirements, note that additional requirements should be added and adapted to the project goals.

Table 19. Example of functional & non-functional requirements[31]

	ID	Requirement	Description
Functional requirements	**F1**	Response provision	The CA should be able to answer user queries. *Example:* • *If the user asks a question such as "What documents do I need to apply for citizenship in Luxembourg", the CA should be able to answer the user input (by triggering the skill/ invocation name that links to the public institution chatbot that has this information).*
	F2	Command execution	The CA should be able to understand and execute commands or requests from the user. *Example:* • *The CA should be able to create an email given an email address from the public institution's bot, call a number or play a music from its own skills.*

[31] Not all will be relevant, and the interoperability case might require additional ones

	ID	Requirement	Description
	F3	Context handling	Ability to handle the conversation context when interacting with multiple systems. The CA should be able to pick up the conversation where it left off when switching between "systems". *Example:* • *If a user asks following questions: "Where can I find tax information for Luxembourg?" and then "What about Slovenia?" the CA should understand the context and provide information about tax.*
	F4	Multilingualism	The CA might support multiple languages to cater to a diverse audience. It should have sound language processing capabilities, including translation and understanding. *Examples:* • *A user in France can ask the CA questions in French, and the CA will understand and respond in French.* • *A user in Germany can also interact with the same CA using German and still receive accurate responses.*
	F5	Error handling	Should an CA not respond, error messages should be issued. Offering a returned response and gauging satisfaction can ensure the system learns from its tested responses. Ability to identify and handle errors effectively, for instance to resolve or reroute the assistants' pathway seamlessly. *Example:* • *If a user makes a speaking error or asks a question that the CA does not understand, it should guide the user to provide a valid input.*
	F6	Monitor	Monitoring ensure everything is working as expected - it is important in maintaining system integrity, performance and accuracy. Metrics help evaluate the system such as response time, resource usage, error rates, error tracking, active users, feature usage, etc. *Example:* • *CA to have analytics available on their developer console to show data such as the number of active users, Intents activated, utterances, common queries, etc.*
Non-functional requirements	**NF1**	Cost-effectiveness	The system should provide optimal functionality and efficiency at a reasonable expense (this threshold should be defined). *Example:* • *The CA should optimize its use of compute and storage resources to keep operational costs to a minimum.*
	NF2	Response Time	The CA should be able to respond to user queries within acceptable time limits. *Example:* • *If a user asks the CA a question, they should receive a response within a few seconds, regardless of whether the question is for the CA or the public institution chatbot.*

	ID	Requirement	Description
	NF3	Performance	Corresponds to how well the CA interacts and if the correct CA or bot responds to the user query. *Example:* • *The user query should be redirected to the most relevant assistant, the CA or the public institution chatbot, to answer it given a certain wake word, invocation name or intent was mentioned.*
	NF4	Security	All data exchange between the CA and public institution chatbot must be secured to maintain privacy and confidentiality. This can be done through encryption to help prevent data breaches. *Example*: • *If a user shares personal data, such as an email address or phone number, the CA should securely manage this data.*
	NF5	Usability	The interoperability connection between the CA and the public institution chatbot should be user-friendly and intuitive for users. *Example:* • *Users should find it straightforward to interact with the CA, with clear prompts for input and easily understandable responses once the user downloads the skill and knows the relevant invocation words.*
	NF6	Consistency across devices	Compatibility of interoperability between the CA and the public institution chatbot to work on different CA devices. *Example:* • *Whether a user is interacting with the Alexa Echo (2nd Gen.) or on Echo Show, the interoperability between the CA and public institution and UX should be consistent and error-free.*

B.2.2 Epics / User Stories overview

Step 4. Define Epics

Epics are large bodies of work that can be broken down into several smaller tasks, namely user stories. Based on the functional and non-functional requirements outlined earlier, related requirements will be grouped into categories, or "epics". Table 20~~Table 22~~ below is an example template of what epics can look like in the context of interoperability.

Table 20. Examples of epics

Epic ID	Epic Name	Related requirements
1	Rich and dynamic user engagement	Response provision, Multilingualism, Context handling, Command execution, Consistency across devices, Response time, Performance, Usability
2	Advances operational intelligence	Monitoring, Error handling
3	Secure and compliant information handling	Security, Error handling

Step 5. Create User stories

User stories help us understand the user perspective. These are created based on user personas, which are fictional representations of main user types, and are assigned to an epic. First, user personas will be defined, then narrative-based scenarios from the perspective of each persona will be written. These personas and stories will help design tests that resemble real-world use of the product.

User personas

User personas represent fictional characters based on actual users and their behaviours, needs, goals, attitudes and pain points. The idea is to use these personas to guide design/test decisions by providing a realistic representation of the key audience that will use the assistants. The use of personas helps in comprehending with the user's needs, facilitating the creation of more user-friendly solutions. Profiles need to be considered also on the different features that we would like to include in the assistants. The personas should consider what topics your assistant covers, which languages it supports and why they would use your assistants. Figure 14~~Figure 12~~ and ~~**Error! Reference source not found.**~~Figure 13**Error! Reference source not found.**~~**Error! Reference source not found.**~~Figure 13~~ below show some examples of general user personas as well as one more detailed example of a user persona.

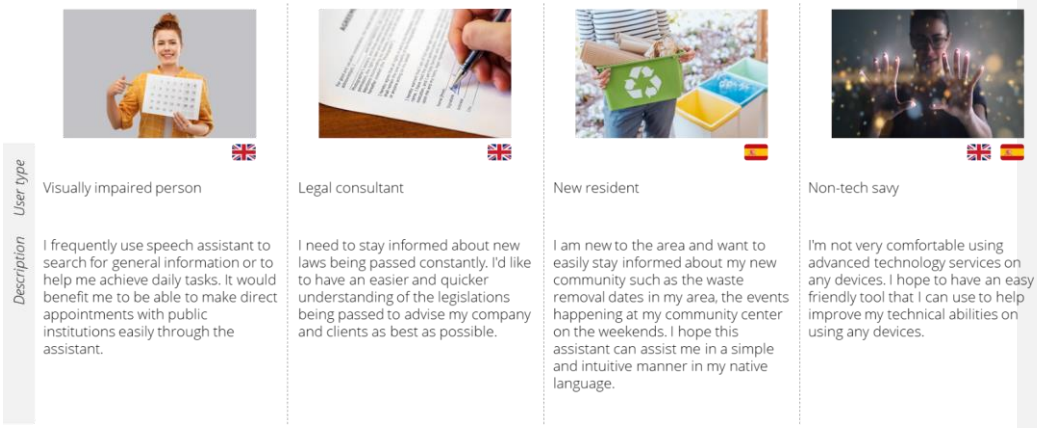


Figure 14. Examples of user personas for different assistants

User stories

User stories help us understand the user perspective. These are created based on user personas, which are fictional representations of our main user types, and are assigned to an epic. After defining user personas, like we have done in the previous section, narrative-based scenarios can be written from the perspective of each persona. These personas and stories will help design tests that resemble real-world use of the product. Table 21~~Table 23~~ below shows an example template of user stories which are based on the user personas created in the previous section.

Table 21. Examples of user stories

Epic ID	Epic Name	US ID	User story

1	Rich and dynamic user engagement	1.1	As a visually impaired person, I need the assistant to execute commands accurately when I want to make an appointment with my townhall to renew my passport.
		1.2	As a new resident, I need the assistant to support my native language, Spanish, when I ask about the waste removal dates or events happening in my community.
2	Advances operational intelligence	2.1	As a visually impaired person, I need the assistant to handle context to provide accurate replies based on my previous queries.
		2.2	As a legal consultant, if there are errors during the chat, I need the bot to handle them effectively to avoid wasting precious time.
3	Secure and compliant information handling	3.1	As a legal consultant, I need trustworthy and up-to date information about legislative documents and my personal data to be handled securely and without any breaches.
		3.2	As a Non-tech-savvy User, I need assurance that my personal details are not shared or misused when I ask about social welfare programs.

B3. Implementation framework: Templates for Phase C – Testing

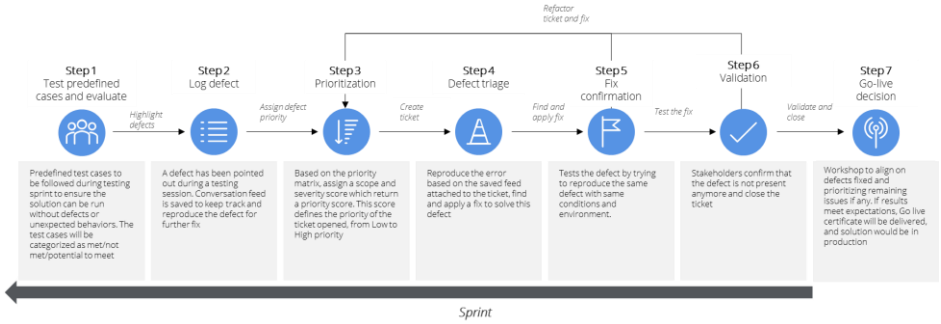


Figure 15. Steps to conduct testing

<u>Zoom in Prioritization (step 3)</u>

To optimize the UAT process, create a prioritization matrix. The matrix will serve as a tool to identify which requirements are of most importance and need to be tested first. This is usually based on the business value, risk, complexity, and impact of each requirement. An example template of such a prioritization matrix can be seen below.

Severity – Does the criteria affect sensitive features of the interoperability?

Scope – Does the criteria affect all users or a minor part of users?

Priority Assessment	
1-2 – High	• To be fixed in sprint 1 • Requires notification to project sponsor & project manager • Does not meet go live criteria if any high priority defect is open
3-4 – Moderate	• To be fixed in sprint 2 • Requires notification to project manager • Does not meet go live criteria if any high moderate defect is open
6-9 – Low	• To be scheduled when time is available or must be scheduled for further enhancements phase • Does meet go live criteria if any low priority defect is open

Table 22. Example of prioritization matrix

		Priority		
	Level	High	Medium	Low
Scope	Large	1 User Story 3.1 & 3.2	2 User Story 1.1	3 User Story 1.2
	Moderate	2 User Story 2.2	4 User Story 2.1	6
	Small	3	6	9

B4. Implementation framework: Templates for Phase D – Deployment & monitoring

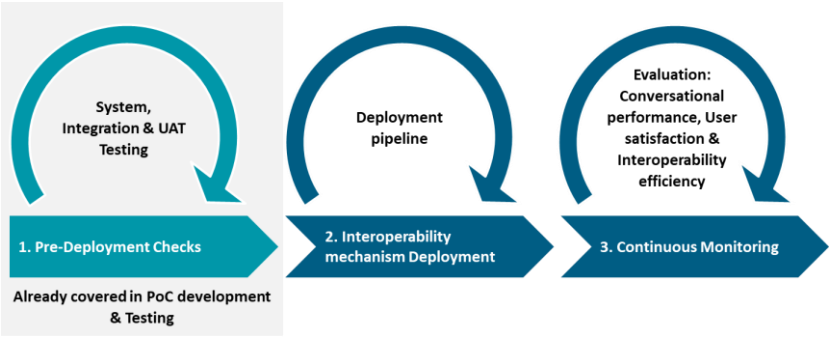


Figure 16. Deployment & Monitoring overview

Define KPIs

Monitoring revolves around observing the interoperability process for any issues that might arise.

KPIs are a critical component in assessing the performance and success of the assistant's interoperability in accordance with the set goals and objectives. These measurable values offer an insight into the effectiveness of the assistant's functionality, UI, interoperability driver, and user satisfaction, aiding in the optimization and improvement of future interactions. Widely available CAs usually have analytics dashboard available for to monitor tools or skill that have been created through their SDKs. The dashboard can display metrics such as customers per day/average, intents used, utterances, retention etc.

The following KPI's present with examples of measures that should be defined before the development of the solution and that could be used to create an internal monitoring dashboard. They can be classified in three categories, the first one focuses on the conversational efficiency of the bot, the second one centres on measures around user satisfaction and retention and finally the last one emphasis on interoperability efficiency between the bots.

Table 23. Monitoring KPIs

	KPI	Explanation	Example / Measure
1. Conversational efficiency provided by the bot	1. Error Rate	The number of errors experienced per a certain number of interactions or a specific time.	The error rate can be measured in percentage, i.e., less than 10%.
	2. Accuracy rate	The extent to which an assistant correctly understands and appropriately responds to user queries.	On a sample of 100 queries, the assistant provided correct and relevant responses to 81 queries. It has an accuracy rate of 81%.
2. User satisfaction	3. User Retention Rate	The percentage of users who return for successive interactions after their first use. Higher retention rates generally indicate positive UXs.	The user retention rate can be measured in percentage, i.e., 85%.
3. Interoperability efficiency	4. Connection Response Time	How long it takes from the moment a question leaves the CA to when an answer from another a public institution chatbot is received. The speed is impacted by the interoperability mechanism on which bot should answer the user depending on the invocation name.	Ideally, connections should take under 10 seconds. Connections beyond 10-15 seconds may be deemed failures, prompting the driver to query another bot or inform the user about the inability to answer.
	5. API Trigger Time	The time taken by the AI driver to trigger an API. Shorter trigger times indicate better system performance.	This should be a short as possible, for example a general good response time is 0.1-1 second.

	KPI	Explanation	Example / Measure
	6. API Failure Rate	The percentage of attempted API triggers that fail. Lower rates are indicative of better system health.	This should be as low as possible, for example 5%.
	7. Successful Connection Rate	The ratio of successful connections from CA to public institution chatbot (requests answered within the expected time) to the total number of requests.	The successful connection rate can be measured in percentage, i.e., at least 80%.
	8. Outbound Connection Requests	A tally of the number of requests sent to other bots helps in measuring the workload on other bots (also gives implicit information on what users tend to ask more).	CA sent 6 out of 12 requests received that day to the public institution chatbot. Showing popularity of skill.

B5. PoC Technical documentation