Cost effective, Multilingual, Privacy-driven voice-enabled Services

www.compriseh2020.eu

Call: H2020-ICT-2018-2020
Topic: ICT-29-2018
Type of action: RIA
Grant agreement Nº: 825081

WP Nº5: Cloud-based platform for multilingual voice interaction

Deliverable Nº5.1b: Updated data protection and GDPR requirements

Lead partner: ROOT

Version Nº: 1.0

Date: 31/05/2021

Document information	
Deliverable Nº and title:	D5.1b – Updated data protection and GDPR requirements
Version Nº:	1.0
Lead beneficiary:	ROOT
Author(s):	Álvaro Moretón, Ariadna Jaramilo, Conrado Castillo, Laura Merlo (ROOT)
Reviewers:	Youssef Ridene (NETF), Thomas Kleinbauer (USAAR).
Submission date:	31/05/2021
Due date:	31/05/2021
Type[1]:	R
Dissemination level[2]:	PU

Document history			
Date	Version	Author(s)	Comments
10/05/2021	0.1	Álvaro Moretón, Ariadna Jaramillo Conrado Castillo, Laura Merlo	First draft of the document
24/05/2021	0.2	Álvaro Moretón, Ariadna Jaramillo Conrado Castillo, Laura Merlo	Final version based on the reviewers' comments
31/05/2021	1.0	Emmanuel Vincent & Akira Campbell	Final version reviewed by the Coordinator and the project manager

[1] **R**: Report, **DEC:** Websites, patent filling, videos; **DEM:** Demonstrator, pilot, prototype; **ORDP:** Open Research Data Pilot; **ETHICS:** Ethics requirement. **OTHER:** Software Tools

[2] **PU:** Public**; CO:** Confidential, only for members of the consortium (including the Commission Services)

Document summary

This document comes with Deliverable D5.5 “Final platform demonstrator and updated data protection and GDPR requirements” (submitted to the European Commission on May 31, 2021). It contains an updated version of the contents of Deliverable D5.1 “Data protection and GDPR requirements” (submitted to the European Commission on May 30, 2019), which provided a comprehensible summary of the main aspects related to the General Data Protection Regulation (GDPR), including an analysis of the type of data involved in voice interactions and the particularities of the processing of personal data in voice-enabled systems where different technologies, such as machine learning (ML), are integrated.

Since the completion and publication of Deliverable D5.1, new guidelines and works on the processing of personal data by voice technologies have been published, which have been considered to extend the existing recommendations on GDPR compliance. The recommendations included in this document follow a more practical approach than its predecessor, focusing on how to enable data subjects to exercise their rights and GDPR compliance in the exploitation state of COMPRISE.

Finally, other aspects such as ethics and cybersecurity are also addressed in this document.

Table of contents

1 Introduction

Deliverable D5.1 “Data protection and GDPR requirements” (submitted to the European Commission on May 30, 2019) provided a comprehensible summary of the main aspects of the GDPR, considering and analysing the particularities of voice interaction technologies and the COMPRISE project, identifying barriers and requirements to comply with the Regulation.

The current document extends the work done in Deliverable D5.1. It analyses new guidelines and papers, good practices carried out by voice technology companies, and relevant changes in the current legislation interpretation. It provides practical recommendations for GDPR compliance of personal data processing operations involving voice technologies and extends such recommendations to the COMPRISE platform and its exploitation.

This document comes with the fifth deliverable of WP5 “Cloud-based platform for multilingual voice interaction”, which brings together WP2, WP3 and WP4 to develop a cloud-based platform that collects anonymised speech and text data from the users and curates it. WP5 aims to provide access to the speech-to-text, spoken language understanding, and dialog management models trained on these data as a service via a web service API. It should be noted that, initially, the content of this document was intended to be integrated into Deliverable D5.5 “Final platform demonstrator and updated data protection and GDPR requirements” (submitted to the European Commission on May 31, 2021). However, due to its length, it has been decided to make a separate document and summarise the content in Deliverable D5.5.

Deliverable D5.1 also provided a series of general recommendations on the implementation of the GDPR, aimed to address GDPR compliance during the development stage of the project. Furthermore, it focused on possible issues that may arise when machine learning techniques are involved in personal data processing.

This document provides new recommendations to multiple stakeholders (e.g., voice assistant designers, voice app developers, user companies, etc) that collect and process personal data through voice technologies on how to implement GDPR requirements. It is more specific and follows a more practical approach (i.e., providing practical solutions that enable compliance data processing through different voice systems).

It also includes additional recommendations related to the exploitation of COMPRISE.

It should be noted that the application — or not — of the GDPR to COMPRISE will depend on the effectiveness of the anonymisation technique applied to the speech and text data collected from the users. This will impact the risk of re-identification (i.e., tolerable or not tolerable) after applying anonymisation, which the data controller should assess to make a final decision.

Lastly, the analyses and recommendations provided have been extended to areas such as ethics and cybersecurity, both critical for the exploitation of voice technologies.

2 Sources analysed

Since the publication of Deliverable D5.1, different materials aimed at providing guidance on how to interpret the law and implement measures to comply with the GDPR have been published. Furthermore, some companies providing voice-related services have

implemented novel organisational and technical measures to improve privacy assurance, especially after the privacy scandal involving big tech company employees or subcontractors listening to voice assistant recordings to carry out human data annotation.

The main sources of information analysed to prepare this document are detailed below:

- **Supervisory Authorities' guidelines:** Several data protection Supervisory Authorities have published guidelines, opinions and/or articles on various topics related to voice technologies (e.g., how to comply with GDPR requirements when personal data processing is carried out through voice technologies or when machine learning is involved in the processing; how to implement privacy by design principle in products, services and processes, etc.). However, opinions, recommendations, and administrative agencies are considered "soft law", i.e., they are not binding, although they increase legal certainty as the regulator clarifies how it interprets the legislation. The following table lists some of the main documents released by Supervisory Authorities, which readers may find useful:

Title	Link	Authority	Description
White Paper: On the record	Click here	CNIL	It aims to present various legal, technical and/or ethical issues and respond to the concerns of voice assistant manufacturers, distributors, and users. It offers advice and guidance to help ensure that voice-based systems and solutions are developed in a way that respects the fundamental rights of the users.
Guidelines 02/2021 on Virtual Voice Assistants	Currently a draft document released for public consultation. See COMPRISE feedback on the document in Appendix B.	EDPB	The EDPB published draft guidelines open to the public for feedback until April 23, 2021. The final version will be available in the upcoming months after the feedback received is considered. These guidelines identify some of the most relevant compliance challenges for virtual voice assistants and provide recommendations to relevant stakeholders on how to address them.
TechDispatch #1: Smart Speakers and Virtual Assistants	Click here	EDPS	This guideline explains what a smart speaker and a virtual assistant is and addresses related privacy issues.
RGPD compliance of processing that em-	Click here	AEPD	This guideline aims to be the first survey for GDPR compliance of products and services that embed Artificial Intelligence components

beds Artificial Intelligence. An introduction			
Guidance on AI and data protection	Click here	ICO	This guideline aims to help organisations mitigate the risks specifically arising from AI following a data protection perspective. It explains how data protection applies to AI projects without losing sight of the benefits such projects can deliver.
Guidelines 4/2019 on Article 25 Data Protection by Design and by Default	Click here	EDPB	These guidelines provide general guidance on the obligation of Data Protection by Design and by Default outlined in Article 25 of the GDPR.
A Guide to Privacy by Design	Click here	AEPD	This guideline provides a methodological focus centred on risk management and accountability that allows organisations and individuals (e.g., developers) to determine privacy requirements by means of practices, procedures and tools

- **Other documents:** Since the publication of Deliverable D5.1, additional documents, such as journals, articles, and reports that analyse aspects related to voice technologies, cybersecurity, machine learning, and ethics, have been published. The examination of these documents has been useful to extract conclusions and recommendations on how to process personal data through voice technologies in a compliant, secure and ethical way. The publications consulted to prepare this document can be found in Section 12.
- **Good practices:** The observation of good practices implemented by voice technology companies should also be considered when implementing GDPR requirements. However, every personal data processing operation should be analysed individually, and the measures implemented adapted following a case-by-case approach to make them as effective as possible.

3 Personal data processed by voice technologies

Section 4.2.1. of Deliverable D5.1 provided an analysis of the concept of personal data, which, according to Article 4.1 of the GDPR "*is any information relating to an identified or identifiable natural person".* Additional topics, such as scenarios where the voice could be deemed an identifier or where the voice signal, its content and the information derived from it could be considered personal data if it is possible to identify an individual through it, were also explained in Deliverable D5.1.

Regarding identifiers, it should be mentioned that when they are used (e.g., User ID Tracking in which natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols), all information (e.g., spoken message)

linked to them is considered personal data, even if such information would have been an identifier on its own.

There, it is important to not only assess personal identifiers that will be revealed in the speech (e.g., name or credit card number within the spoken message) but also whether all the speech (i.e., any information) is linked — or not — to an individual via the use of technical identifiers or trackers.

If an identifier (e.g., serial number of the speech assistant device, or client ID) is used to single out (i.e., identify) the user when it connects (so each time it would be possible to know which user it is, even if their name is not known), then all the information shared, the content of the speech — even that revealing simple preferences or interests — or words, is to be considered as personal data.

4 Good practices to implement GDPR principles

This section provides a series of recommendations on measures and good practices to be implemented to comply with each of the GDPR principles that were introduced in Deliverable D5.1. The recommendations focuses on the exploitation of voice technologies and, more specifically, of the COMPRISE solution.

4.1 Lawfulness, fairness, and transparency

According to the "lawfulness, fairness and transparency principle" (Article 5.1A) of the GDPR, "*personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject*". This principle was explained in detail in Section 4.3.1.1 of Deliverable D5.1. The implementation of all the elements of the lawfulness, fairness and transparency principle is analysed separately in the following subsections.

4.1.1 Lawfulness

The legal grounds for the lawful processing of personal data are set in Article 6 of the GDPR.

Article 5.3 of the e-Privacy Directive, on the other hand, establishes that actors who wish to store or access information stored in the terminal equipment (voice assistants are considered terminal equipment) of a subscriber or user in the EEA, requires the end user's previous consent. However, the same article also establishes that the end user's consent is unnecessary when the personal data processing is "*strictly necessary to provide an information society service explicitly requested by the subscriber or user*"[3]. Consequently, when personal data is processed through voice technologies, the purpose of the processing should be analysed to detect if consent is an adequate legal ground or, on the contrary, the processing should be based on another legal basis.

When voice technologies process personal data, it is necessary to consider the purposes and circumstances of the particular data processing operation, as the legal bases for the processing may change depending on them.

[3] European Data Protection Board. (2021). “Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0”.

Below are listed some examples of personal data processing operations through voice technologies and the different possible legal bases that may apply to each of them, in accordance with the CNIL and EDPB guidelines:

- If the processing of personal data is necessary to provide a service expressly requested by the user, consent is not necessary. Therefore, the legal basis could be the execution of a contract to which the data subject is a party (Article 6 (1.b) of the GDPR). This could be the case of a generalist voice assistant that aims to respond simply and quickly to recurring functional needs (e.g., searches on the internet)[4].
- If the processing of personal data (e.g., commands history) is intended to create or enrich a profile to improve the services provided to the client or for advertising purposes, it should be considered to have a different purpose than providing basic services. Therefore, it should be treated as a completely different processing, and its legal basis should be analysed separately.
- If personal data is processed for personalisation purposes strictly necessary for the provision of the service expressly requested by the user, the legal basis could be the execution of a contract to which the data subject is a party (Article 6 (1.b) of the GDPR). However, a case-by-case approach should always be applied.
- If the processing is intended to enrich or create an advertising profile, by nature not strictly necessary for the provision of the service requested by the user, then the user's consent must be obtained for this particular purpose[5].
- Some voice assistants offer users the option to identify themselves via their voices to access services that might differ for each of them. This functionality aims at the unique recognition of the user through the biometric processing of their voice (extracting samples of the voiceprint), which is considered a processing of sensitive information in accordance with Article 9 of the GDPR (see Section 4.2.2. of Deliverable D5.1) and requires enhanced protection. Processing of special categories of personal data is prohibited unless it's done under one of the legal bases contemplated in Article 9 (e.g., obtaining explicit consent from the data user). It is recommended to ensure that the processing of biometric data is deactivated by default and conditional on the explicit consent of each person whose voice is likely to be processed this way. The voice recognition function should only be activated at the user's initiative and not through a permanent analysis of the voices heard by the assistant (e.g., the user family members, visitors, etc.). Additionally, to comply with the requirements established by Article 7 of the GDPR for valid consent, the controller should offer an alternative identification method to biometrics, so the consent is free[6].
- Data controllers processing personal data to provide services through voice apps or skills should also choose their own legal bases for the processing of personal data carried out for its own purposes (e.g., the processing of personal data through a voice app aimed at managing a bank account could be carried out under the legal basis of the execution of a contract with the bank).

[4] Commission Nationale de l'Informatique et des Libertés. (2020). ” Exploring the ethical, technical and legal issues of voice assistants”. Retrieved from: https://www.cnil.fr/sites/default/files/atoms/files/cnil_white-paper-on_the_record.pdf

[5] European Data Protection Board. (2021). “Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0”.

[6] European Data Protection Board. (2021). “Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0”.

- A voice-enabled device installed in a private property (e.g., a house), whose operation does not imply the transmission of data outside of said environment (i.e., the assistant performs all operations locally, without requiring an exchange with a remote server nor the transmission of data to a data controller) can potentially benefit from the exemption provided in Article 2 (2.c) of the GDPR. Devices that are not logged in and only allow the user to start or stop household electrical equipment, like voice assistants, would escape the application of the GDPR[7].

There may be cases where, though the data controller intends to obtain the user's consent as the legal ground for a processing operation, the voice-enabled system (e.g., a voice assistant) accidentally records the voice of a third person who is not intending to interact with the device (i.e., someone different from the registered user). Below are described a couple of examples that illustrate this scenario:

- A voice assistant does not use a voice recognition function, and the wake-up word is not only the same for everybody but publicly known (e.g., "Hey--- ","hello…"). In this scenario, any person — different from the user — that has the account could interact with the assistant and have their data processed without consent.
- A third party's voice is recorded as part of a background conversation when the user with the account is interacting with the assistant (e.g., the user is asking something to the assistant and the assistant also records the voice of the user's wife that is speaking at the phone at the same time).
- The assistant wakes up by accident and starts recording.

When data controllers identify accidental recordings of third parties, and there is no legal ground to process such data, as it is considered that consent hasn't been provided, the recordings should be immediately erased[8]. It seems reasonable to recommend a proactive attitude from data controllers and processors, who must track (automatically or manually) the recordings collected to detect and eliminate those accidentally obtained.

COMPRISE
In the case of COMPRISE, either the anonymisation effectiveness or the level of anonymisation configured by developers should be analysed. If the anonymisation reached is enough so the data collected through the voice app can be considered non-personal data, there would be no need to carry out the processing under one of the legal grounds of Article 6, as the GDPR wouldn't apply. On the contrary, if the data is still considered personal data after applying the corresponding anonymisation techniques, the guidelines above should be followed to identify the most accurate legal grounds for the processing.

[7] Commission Nationale de l'Informatique et des Libertés. (2020). "Exploring the ethical, technical and legal issues of voice assistants". Retrieved from: https://www.cnil.fr/sites/default/files/atoms/files/cnil_white-paper-on_the_record.pdf

[8] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".

4.1.2 Transparency

According to the “transparency principle”, data controllers need to inform data subjects, amongst other information, about what personal data is going to be processed and the purposes of the processing (see Section 4.3.1 of Deliverable D5.1).

It is important that users clearly understand what data is processed, how it is processed and for which purpose, as consent won't serve as a legal ground for carrying out a processing operation if not considered “informed”.

Moreover, when the legal ground for the processing is different from consent (e.g., personal data processing is necessary for the execution of a contract between the controller and the data subject), the fulfilment of the transparency principle is also indispensable to comply with the GDPR.

There is a general perception that voice technology companies have failed to adequately inform users about their personal data processing. For example, in 2019, several media published that different voice technology companies hadn’t properly informed their clients about their practice of hiring humans to review clips of conversations between devices and their users. In this sense, it is critical that data controllers processing data through voice technologies properly inform data subjects about the different categories of personal data they collect and process.

Voice assistants and voice apps may collect several types of personal data, not only the user’s voice, as many believe. Said data can be included in different categories, as exemplified below:

- Voice data: Includes the voice signal and the speech content.
    - Several pieces of information could be extracted from the voice signal:
        - General traits of the speaker (e.g., gender, age, ethnic origin, etc.)
        - Mental condition (e.g., stress, relaxation, depression, etc.) and health condition
        - Emotional state (e.g., angriness, happiness, nervousness, etc.)
    - Information could also be extracted from the speech content, such as:
        - Words or utterances explicitly mentioning the user's identity, general traits related to the speaker's background (e.g., age, nationality, etc.), information related to the user’s health, or otherwise critical information (e.g., credit card number, home address, etc.)
        - Information not revealed by the dialogue outcome (e.g., user preferences revealed by asking the system about similar products or a general category of products before settling on one)
        - Private information about other individuals than the registered user that has been recorded (e.g., age of a user’s family member)
        - Search and commands history
        - Device data
        - App metadata (e.g., how metadata is being used)
        - Locations
        - Calendar information
        - User’s contacts
        - Browsing history

Therefore, it follows that, for every scenario, the user should be informed of the different pieces of information that can be extracted from their voice.

Another important aspect that should be properly notified to the data subject is the purpose or purposes of the processing of the personal data processed by the voice system. Below are listed some examples of typical purposes for the processing of personal data recorded through voice assistants or voice apps:

- To respond to a user's requests
- To improve recommendations functionalities
- To improve voice recognition functionalities
- To improve conversations comprehension (NLU) and conversation management
- If the voice system is able to recognise emotions or health conditions from the voice, this should be notified to the user, specifying its purpose (e.g., analysis of the voice to detect the emotional status of the user and adapt advertisement)
- Profiling to personalise answers, services or advertisement
- Interaction with other products apps (e.g., calendar) to provide a more accurate and adapted response
- The collection of data different from the voice could be used for different purposes, for example:
    - In the case of location data, to provide relevant notifications (e.g., closest shops) or more accurate and relevant information (e.g., answer which are the closest restaurants to the user's location)
    - In the case of contact lists, to write emails, make calls, write messages, etc.

Often, voice assistants service providers are global companies that carry out an assortment of activities (e.g., e-commerce, social media, telecommunications). Therefore, data subjects should be adequately informed whether their use of the voice assistant will be linked to other processing activities carried out by the provider (e.g., to enhance user profile)[9].

A voice assistant's ecosystem is quite complex. The roles and identities of those processing personal data usually are not clear. Furthermore, personal data collected through a voice assistant can be processed by several independent data controllers. It is critical to inform the data subject about the different stages of the processing and actors involved[10]. For example, data subjects should be aware of whether an assistant is able to send information to service providers in order for them to execute the desired service (e.g., the name and address will be needed to provide a food delivery service).

The data controller to the personal data collected through a voice app developed for a specific voice assistant (also known as skill) to provide a service (e.g., manage a bank account, ordering food, etc.) should also provide information to the app users on the processing of their personal data. Moreover, these data controllers should have their own privacy policy (easily accessible to the user), containing all the information related to the personal data processing operation.

According to the research "Measuring the Effectiveness of Privacy Policies for Voice Assistant Applications"[11] carried out by Clemson University, most voice apps available

[9] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".
[10] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".
[11] Liao S., Wilson C., Cheng L., Hu H., Deng H. (2020). "Measuring the Effectiveness of Privacy Policies for Voice Assistant Applications". Retrieved from: https://arxiv.org/pdf/2007.14570.pdf

in app stores don't have their own privacy policies (of all the skills analysed for the research, only 17,952 out of the 64,720 (28%) had a privacy policy). Furthermore, several skills have broken or duplicated privacy policies links.

The research data also showed that, many of the skills' privacy policies analysed fail to clearly define what data practices the skill is capable of, as they are too generic.

Moreover, privacy policies provided by voice assistant companies may be complex and lengthy due to the bundling of the voice assistant user account with other services provided by the company, such as email or video streaming. The EDPB recommends avoiding this practice to comply with the transparency principle[12].

One of the main problems to comply with the transparency principle is the multiplicity of users (registered, non-registered, accidental) that have to be informed about their data being processed. Once a smart speaker is up and running, it is unlikely that users other than the person who installed the device will have read the written Privacy Policies that came with it[13]. The CNIL recommends that voice assistants/voice apps directly inform the users about privacy policies[14] or, at least, provide a first layer of information that the user can opt to extend (maybe through a question/answer system) or that invites to read the corresponding privacy policy. Such a system would prevent a prolonged reading of the terms of use and privacy policy and help inform users different from those that created the account. If the system is capable of recognising voices, it should be informed on the first layer of the privacy policy to the new speaker.

Following the same line, the EDPB recommends solving information asymmetries for the different types of users both by making the voice assistant more interactive and informing of the current status of the assistant at any time (e.g., it can be listening locally for the detection of wake-up expressions or interacting with a remote server to resolve a command, recording environmental sounds — including background conversations, or interacting with an unknown user). The EDPB proposes to make the man-machine dialogue more interactive or use specific signals to broadcast the status of the voice assistant (e.g., icons or lights)[15].

Finally, for AI solutions, some parameter should be implemented to ensure transparency and accountability. As explained in Section 2 of Deliverable D5.1, voice-enabled systems are based on machine learning, which is used to understand the user better and manage conversations (e.g., take a decision on which answer should be provided when a user makes a specific request). According to the "GDPR compliance of processing that embeds Artificial Intelligence"[16] guidelines, published by the AEPD, any AI technical solution

[12] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".
[13] European Data Protection Supervisor. (2019). "TechDispatch #1: Smart Speakers and Virtual Assistants". Retrieved from: https://edps.europa.eu/data-protection/our-work/publications/techdispatch/techdispatch-1-smart-speakers-and-virtual_en
[14] Commission Nationale de l'Informatique et des Libertés. (2020). " Exploring the ethical, technical and legal issues of voice assistants". Retrieved from: https://www.cnil.fr/sites/default/files/atoms/files/cnil_white-paper-on_the_record.pdf
[15] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".
[16] Agencia Española de Protección de Datos. (2020). "RGPD compliance of processings that embed Artificial Intelligence An introduction". Retrieved from: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia-en_0.pdf

should provide a certified answer to the following issues to be considered a mature technology, capable of complying with basic requirements of accountability and transparency:

- Accuracy, precision or error rates required by the processing
- Data input quality requirements
- Precision, accuracy or effective error rates of the AI-based solution depending on the appropriate metrics to measure the eligibility of such an AI-based solution
- Convergence of the model when training or evolve the AI-solution
- Consistency in the results of the inference process
- Algorithm predictability
- Any other parameters to assess the AI component

As voice technologies solutions are based on AI and use machine learning techniques to train the models, the issues above should be considered (from the design to the development of the solution) to ensure compliance with the transparency principle.

COMPRISE
In COMPRISE, based on the re-identification test analysis results, it should be decided whether the data collected and uploaded to the platform is treated as personal or non-personal data. If the data uploaded to the platform is considered personal data despite the anonymisation techniques applied, the information mentioned in Article 13 of the GDPR should be provided to the data subject (see Section 4.4.1 of Deliverable D5.1). Also, good practices should be considered as well. In both scenarios (the data collected is considered personal data or the data collected is considered non-personal data after applying anonymisation), it would constitute a good practice to inform about the anonymisation process applied to the speech data collected and stored in the COMPRISE Cloud Platform, as well as of the existing risks after applying anonymisation (i.e., if there is a risk of re-identification). Moreover, it would also be necessary to inform (for both scenarios) that the data collected and stored in the COMPRISE Cloud Platform could be shared, given prior consent of the data subject, with third parties. Of particular importance would be to stress that humans (annotators, developers, administrators) may access the dataset. If COMPRISE apps process any personal data different from speech (location data, identification data, calendar information, etc.), the data subjects should be informed as required in the GDPR. Finally, the information provided regarding the anonymised datasets that are collected through COMPRISE apps and uploaded to the COMPRISE Cloud Platform, whose processing purpose is to train models, should be distinguished from that received from service providers through the application to execute a service (e.g., a food delivery company that need the user's name, telephone and address, provided in the command, to deliver the order). The service provider should have its own privacy policy and process the personal data under a valid legal ground.

4.1.3 Fairness

As explained in Section 4.3.1.1 of Deliverable D5.1, the "fairness principle" seeks a fair treatment of the data subject when their personal data is processed, meaning that discriminatory and arbitrary treatment should be prevented.

One typical form of discrimination associated with voice systems is the inability of voice-enabled systems to recognise different accents with the same precision, also known as "the accent gap" problem. It occurs because audio samples used to train speech-recognition models came from the same background (usually white, male native speakers), resulting in a more accurate understanding of this segment of the population than others that have not been properly represented[17].

A poorly representative dataset may also cause other biases that jeopardise the accuracy of voice technologies, such as gender bias.

COMPRISE
COMPRISE adapts user-independent models trained on anonymised data in the cloud (user's speech is automatically anonymised before being sent to the cloud) on the user's own data. User-independent speech and language models are personalised to each user by running additional computations on the user's device or on a Personal Server. This improves speech recognition accuracy for all users, boosting user's experience and inclusiveness.

4.2 Purpose limitation

According to the "purpose limitation principle" (Article 5.b of the GDPR) (see Section 4.3.1.2 of Deliverable D5.1), personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. Regarding personal data processing in voice systems, the main purposes notified to data subjects are:

- To execute user's requests,
- To train and improve the machine learning model, and
- To recognise the user's voice[18].

However, it is quite common that data controllers use the data collected through voice assistants for different purposes than those notified to the data subjects. For example, profiling to provide unsolicited personalised services like advertising or sentiment and health condition analysis. In this scenario, it would be necessary to specifically inform the data subject about the new purposes and carry them out based on the corresponding legal ground (e.g., obtain consent for the specific purpose).

[17] Moretón A., Jaramillo A. (2021) "The accent gap problem in minorities and dialect speakers". Retrieved from: https://www.technology.org/2021/04/23/the-accent-gap-problem-in-minorities-and-dialect-speakers/

[18] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".

4.3 Minimisation

According to the minimisation principle (Article 5.c of the GDPR) (see Section 4.3.1.3 of Deliverable D5.1), the processing of personal data must be limited to what is necessary to the purposes for which the processing operation is carried out (legitimate purposes). This means that only data relevant and adequate for the purpose or purposes of the processing should be collected. Besides, personal data should only be processed when the purpose of the processing cannot be reasonably fulfilled by other means.

The functions offered by voice assistants require speech processing for keyword detection, automatic transcription and analysis and interpretation of the order. Additionally, if the assistant is connected to an account, only data essential for the operation of the account and the user interactions with the assistant should be processed[19].

Fortunately, voice companies have been implementing different measures to minimise the amount of personal data processed and the impact on the users whose personal data needs to be processed over the past years. Below are listed some examples of measures and good practices implemented by companies offering voice solutions:

- One of the most popular measures voice assistants have implemented is wake word detection. This technology inspects acoustic patterns detected when the wake word has been spoken, using an on-device buffer in the temporary memory (RAM), which is continuously overwritten[20]. No audio is streamed to the cloud until the wake word is detected, meaning that only the speech data necessary to interact with the voice system (e.g., ask something to the voice assistant) is processed in the cloud. However, it is still possible that voice-enabled systems erroneously believe to have heard the wake word when it was never uttered and start recording.
- Some voice assistant companies have recently started to implement the "Guest Mode" option. It enables users to use the assistant without signing a user account, hence offering stronger privacy guarantees, as the assistant won't offer personalised answers or save interactions. The guest mode may be activated via voice, with an icon appearing on the display indicating it has been switched on[21]. While the guest mode limits the interaction, it enables popular features like asking questions, controlling home smart devices or playing music.
- Voice apps can request the user permission to collect different types of information for different purposes (e.g., provide relevant information in responses, complete transactions, etc.).To limit the collection of information to what is strictly necessary for the app to execute its functions, some voice assistant providers include in their app builders an option that allows developers to toggle on only the permissions the app will need, so only these are required to the app user[22].

---

[19] Commission Nationale de l'Informatique et des Libertés. (2020). ” Exploring the ethical, technical and legal issues of voice assistants”. Retrieved from: https://www.cnil.fr/sites/default/files/atoms/files/cnil_white-paper-on_the_record.pdf

[20] “Alexa Privacy and Data Handling Overview”. Retrieved from: https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf

[21] Wiggers K. (2021). “Google launches privacy-sensitive Guest Mode on Google Assistant devices”. Retrieved from: https://venturebeat.com/2021/01/13/google-launches-privacy-sensitive-guest-mode-on-google-assistant-devices/

[22] Amazon. Configure Permissions for Customer Information in Your Skill. Retrieved from: https://developer.amazon.com/es-ES/docs/alexa/custom-skills/configure-permissions-for-customer-information-in-your-skill.html

- Voice assistant providers have implemented protocols for sharing only very limited information with app developers creating skills or apps for them (e.g., not sharing voice recordings with skills developers). For example, not sharing personal data with third parties if the customer has not specifically requested it or provided permission (e.g., by using a similar framework to mobile providers)[23].
- Cloud independent voice assistants running offline have been created, so no personal data from users is collected and stored. Only SLU components are trained on servers when inference takes place on the device[24].

Another good practice to comply with the minimisation and privacy by default principles is to activate the app's privacy-preserving configuration options by default. Users should control and decide whether they want to modify this initial configuration sacrificing some privacy in exchange for utility or a more personalised service. Moreover, it should be properly informed and warned about the consequences and risks arising from their decision (e.g., personal data will be shared with third parties). This good practice is not limited to voice apps but serves all apps collecting personal data.

The AEPD provides some examples of operations that could be included in the privacy panel and configured by users. As indicated before, the most privacy-preserving options should be activated by default, allowing the user to change them if it considers it appropriate[25]. Appendix A lists these examples.

COMPRISE
COMPRISE is a privacy by design solution that enables the creation of voice apps compliant with the minimisation principle. COMPRISE does not process data in the cloud for STT, NLU, or TTS, since everything is processed on the user's device or on a Personal Server. The developer has included all the elements needed for this kind of processing directly on its app. COMPRISE only stores data in the cloud and uses it for training the models used by the app. Additionally, COMPRISE focuses on ensuring privacy in the training branch. To do so, two innovations that complement each other have been introduced: a new privacy-driven speech transformation (COMPRISE Voice Transformer) and a new privacy-driven text transformation (COMPRISE Text Transformer). The privacy-driven speech transformation is applied to the citizen's speech signal before it is sent to the cloud to learn large-scale user-independent speech-to-text models from the speech data gathered from all users. The proposed transformation will result in a new anonymised speech signal from which sensitive attributes related to the user's

[23] Alexa Privacy and Data Handling Overview. Retrieved from: https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf

[24] Coucke A., Bluche T., Doumouro C., Lavril T., Saade A., Caulier A., Gisselbrecht T., Primet M., Ball A., Leroy D., Caltagirone F., Dureau J. (2018). "Snips Voice Platform: an embedded Spoken Language Understanding system for private-by-design voice interfaces". Retrieved from: https://arxiv.org/pdf/1805.10190.pdf

[25] Agencia Española de Protección de Datos. (2020). "Guidelines for Data Protection by Default". Retrieved from: https://www.aepd.es/sites/default/files/2020-10/guia-proteccion-datos-por-defecto-en.pdf

identity have been removed while keeping enough information to train a speech-to-text tool.

The privacy-driven text transformation is applied to the text before it is sent to the cloud and addresses two tasks: identifying the parts of the text to be transformed and performing the actual transformation into an anonymised text. This is done by replacing words and expressions carrying personal information with random alternatives while preserving the sentence structure.

Besides implementing the above-mentioned privacy-driven measures enabled by COMPRISE, app developers and voice assistant providers using the solution should consider the implementation and activation by default of other privacy preservation measures (see Section 4.3 on Minimisation).

Lastly, data controllers using COMPRISE should consider the possibility of granting control of the privacy level offered by the app to the user through a configuration privacy panel. For example, selecting which type of personal information should be removed from their speech, like special categories of personal data (i.e., sensitive information) or any information considered private. In this regard, it would be critical to inform the user of the consequence and risks arising from the privacy configuration chosen.

4.4 Storage limitation

According to the "storage limitation" principle, personal data should be kept in a form that permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed (see Section 4.3.1.5 of Deliverable D5.1).

Currently, several voice assistants offer users the option to delete their personal data whenever they want. This option, however, is not always fully effective as, often, devices only allow users to delete certain types of data (e.g., some devices only allow deleting voice data). Furthermore, handing over the task of deleting personal data to the user contradicts the principle of storage limitation, according to which data controllers must ensure that data is not kept for longer than is genuinely necessary for the purpose of the processing[26].

An automatic deleting system that periodically removes personal data collected could be implemented to tackle the issue above. On this matter, the CNIL recommends determining different retention periods, depending on the type of data collected (e.g., the data associated with the user account can be kept longer than ad hoc requests made with the vocal assistant).

As for personal data that has been accidentally collected, data controllers must verify whether a legal basis for processing this data exists. If there is none, data should be immediately deleted[27].

[26] European Data Protection Board. (2021). “Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0”.
[27] European Data Protection Board. (2021). “Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0”.

COMPRISE
In the case of COMPRISE, if after carrying out an assessment, the anonymisation is considered to be fully effective, and the identification of the data subject is no longer possible, the data storage principle won't apply. On the contrary, if the anonymisation is not considered fully effective after the assessment, and there is a non-tolerable risk of re-identifying the data subjects, appropriate measures should be applied to fulfil the data storage principle.

4.5 Integrity and confidentiality

According to this principle, personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and accidental loss, destruction, or damage.

Cybersecurity measures are explained in detail in Section 9.

5 Data subjects' rights

The GDPR provides data subjects with rights to ensure the control and protection of their personal data. Data controllers have to implement protocols to fulfil the data subjects' requirements when exercising their rights, without undue delay (see Section 4.4 of Deliverable D5.1).

The mechanisms that allow data subjects to exercise their rights should be adapted to the user interface of the technology through which personal data is being processed. In the case of voice assistants, it is recommended to enable data subjects to exercise their rights via voice (aside from classical methods like interaction through the screen). Data subjects (registered or non-registered users) should also be able to exercise their rights of access, removal (e.g., if they have withdrawn their consent), restriction of processing, and portability (under the conditions established by the GDPR) electronically and in a simple way[28]. For this purpose, the provision of self-service tools will be a great option.

Below are listed some of the functionalities that voice assistants already provide to facilitate users the exercise of data subjects' rights[29]:

- Users can review their voice recordings at any moment (i.e., right to access), which should be appropriately ordered and easily identifiable.
- Users can access the list of their voice recordings and delete them (i.e., right to erasure). Some voice assistants provide different options, such as deleting one-day recordings or directly removing all recordings stored.
- Users can decide how long their voice recordings are saved/stored before being deleted.
- Users can opt out from having their voice recordings used by voice assistant providers for specific purposes, such as offering services, creating new features, or improving speech interaction.

[28] Commission Nationale de l'Informatique et des Libertés. (2020). " Exploring the ethical, technical and legal issues of voice assistants". Retrieved from: https://www.cnil.fr/sites/default/files/atoms/files/cnil_white-paper-on_the_record.pdf

[29] Skinner C. (2021). "Every Alexa privacy setting and how to change them". Retrieved from: https://www.techradar.com/how-to/every-alexa-privacy-setting-and-how-to-change-them

- Users can revoke permission from third parties to access the data subject's personal data.

Data controllers should immediately remove the data subject's personal data when data subjects exercise their right to revoke consent or erasure. However, a re-identification risk may remain in some machine learning models. Data controllers should employ models that don't limit their ability to stop processing data subject's personal data when required and implement measures (e.g., anonymisation) to mitigate re-identification risks[30].

COMPRISE
In the case of COMPRISE voice apps, the re-identification risk should be assessed after applying the corresponding anonymisation technique. If the resulting risk is not tolerable, it would be highly recommended to treat datasets as personal data. It is possible that, after applying the corresponding anonymisation technique, the data controller finds itself incapable of identifying an individual within the anonymised set only by resorting to the dataset's information. Nonetheless, it would still be possible that, from the assessment of the anonymisation technique employed, the data controller concludes that there is still a considerable risk of re-identification by linkage. In this scenario, the data controller can choose to manage the anonymised dataset as personal data to comply with the GDPR requirements. But how is the controller going to fulfil its obligations toward data subjects' rights if it is no longer possible to identify them in the anonymised datasets? The GDPR provides a possible solution for these cases in Article 11: 1. *If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.* 2. *Where, in cases referred to in paragraph 1 of this Article, the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 20 shall not apply except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification.* This article emphasises that GDPR requirements should not be used as an excuse for processing more data than necessary and limits some of the data controller obligations.

6 Data controller and data processor

Deliverable D5.1 explained the concepts of data controller and data processor (see Section 4.2.6 for the definition of ‘data controller’ and Section 4.2.7 for the definition of ‘data

[30] European Data Protection Board. (2021). “Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0”.

processor') and provided an overview of their main responsibilities (Section 4.5). This section provides additional guidelines on how to identify controllers and processors of personal data collected through voice technologies.

6.1 Identification of the data controller and the data processor

Before evaluating the aspects that should be considered when identifying data controllers and data processors involved in the processing of personal data collected through voice technologies, the EDPB's guideline on the concepts of controller and processor in the GDPR will be briefly analysed.

Below is a summary of the EDPB's "Guidelines 07/2020 on the concepts of controller and processor in the GDPR"[31], which brings the most recent criteria on the aspects to consider when intending to identify data controllers and processors.

- The data controller is the entity which decides on the purposes ("why") and means ("how") of the personal data processing operations, while the data processor processes personal data on its behalf following its instructions.
- Sometimes, the controllership may be defined by law, i.e., a specific law determines the personal data processing operation(s) and which entity is the controller of the operation or the criteria to nominate the controller.
- Some contracts may specify which parties are the data controller and the data processor of the personal data processing operations. Even when the contract is silent on this matter, it is possible to determine which of the parties is the controller and the processor by analysing the terms agreed between them. This is particularly useful in complex environments where innovative information technologies are used and where the different actors tend to see themselves as "facilitators" and not responsible for the personal data processed.
- In the absence of legal provisions or contractual specifications, the qualification of a party as a controller and processor can be determined on the basis of an assessment of the factual circumstances surrounding the processing, such as:
    - Analysing and trying to determine why the processing is taking place and the role of the different actors involved. For example, one company (the controller) determines the purpose (why) and means (how) of the processing of the personal data, while another (processor) simply processes the data following the instructions of the first. The discretion of these actors should be assessed when determining purposes, i.e., the freedom and autonomy of the party taking the decisions when determining the purposes of the personal data processing.
    - Assessing which entity is deciding on the "means" (how) of the processing. These "means" are not limited to technical ways of processing data but involve additional organisational elements such as "which data shall be processed?", "for how long?", "who shall access the data?" etc. It should be noted that, in some cases, technical and organisational matters may be decided by the data processor, but the essential elements of the decision (e.g., "for how long", "who will have access", or other questions that are

[31] European Data Protection Board. (2020). "Guidelines 07/2020 on the concepts of controller and processor in the GDPR". Retrieved from: https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202007_controllerprocessor_en.pdf

essential to the core of lawfulness of processing), remain inherent to the controller.

The Article 29 Working Party (now European Data Protection Board –EDPB) "Opinion 8/2014 on Recent Developments on the Internet of Things"[32] also provides some principles and recommendations for identifying data controllers, focusing on three specific IoT developments: wearable computing, quantified self, and domotics. However, as stated in the Opinion, these recommendations and principles may apply outside its strict scope and cover other developments. As voice-enabled devices fall into the category of IoT, these principles should be considered when determining which entity is the data controller of the personal data processing operation.

According to Opinion 8/2014, the following entities could be considered as controllers:

- **Device manufacturers**: When they also develop or modify the device's operating system or install the software determining its functionalities (when and to whom data will be transmitted, for which purposes, etc.). For example, a smart speaker manufacturing company that commercialises the device, integrates the voice assistant software in it, and decides on the purposes and means of the processing of the personal data collected through it.
- **Software developers of third-party applications**: When the data subject's data collected through IoT sensors and stored by the IoT device manufacturer can be accessed and processed. For example, a developer that creates voice apps or skills for a particular voice assistant and processes personal data collected through it for his/her own purposes.
- **Other third parties**: For example, a company or another entity (e.g., a hospital) that has integrated a voice-enabled system into its daily activities (e.g., patients' registration and description of their symptoms) and uses the personal data collected for its own purposes.
- **IoT Platforms**: As third parties, IoT platforms have no control over the type of data collected through IoT devices. However, they could be qualified as controllers for processing operations where they collect, and store data generated by the IoT device if they have determined the purposes for which such data is processed.

Unfortunately, each of the profiles indicated before may take one or several roles (controller, joint controller, processor) for a single data processing operation whereas carrying out another role for another data processing. Hence, adopting a case-by-case approach when analysing scenarios is critical. For example, a voice assistant designer may act as a data controller when determining the purposes and means of a processing but may intervene as a data processor when processing personal data on behalf of the app developer[33].

In this respect, the CNIL has analysed some typical examples of personal data processing through voice assistants and voice apps, identifying the data controller for each of them. The conclusions of the CNIL seem to align with the principles exposed before.

The table below summarises the different scenarios considered by the CNIL.

[32] Article 29 Data Protection Working Party. (2014). "Opinion 8/2014 on the Recent Developments on the Internet of Things". Retrieved from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf

[33] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".

Purpose of data processing	Controller / Processor
General uses (e.g., general requests) of a generalist voice assistant (e.g., ask to search for something on the internet)	Here, the assistant designer is responsible for the processing to the extent that it determines the purposes (provision of the voice assistance service) and the means (processing through the assistant linked to a user account).
Specific service (i.e., skill) provided through an app that uses the voice assistant (e.g., for managing a bank account)	Here, the app developer is responsible for the processing related to the provision of the service since it determines the purposes and essential means of the processing associated with the request, allowing interaction with the assistant. (e.g., the app developer offers a dedicated application that allows the user, customer of the bank, to manage their accounts remotely). Furthermore, it decides on the processing means even if the processor (the assistant designer) plays an important role in determining them. When the user interacts with the assistant, their voice goes through the servers of the assistant's designer to be transcribed as text and interpreted, meaning that the bank's response is recorded in the information system of the assistant's designer to be synthesised. Therefore, the latter can access the information that circulates through its servers to answer the questions issued by the user.
The enhancement of the service by improving voice assistant's functions. This may mean having better visibility into the device's uses by implementing usage and operation statistics and correcting keyword detection capabilities, automatic speech recognition, and natural language understanding. Artificial intelligence systems integrated into voice assistants require data to be trained. This may also involve human supervision to improve system's performance.	In these cases, although the purpose of improving the service may lead to the processing of data resulting from the use of applications provided by third parties, there is only one data controller: the designer of the voice assistant.

COMPRISE

In the case of COMPRISE, different considerations should be addressed to identify the roles of data controllers and processors to the data collected through the COMPRISE apps.

As previously mentioned, the re-identification risk associated with the personal data collected through the COMPRISE apps and uploaded to the platform should be assessed to decide whether it will be treated as personal data or non-personal data. This is essential to determine the responsibilities of the different parties involved. Additionally, it is possible that some of the parties involved in the processing of data collected through COMPRISE apps need to access some of the personal data of the user, even when the data uploaded to the cloud is anonymised (e.g., a service provider that has developed a voice app but needs some personal data from the user to execute the service such as name, telephone and address to complete a delivery).

COMPRISE solutions can be exploited in different ways. They often involve different parties processing the data collected through the COMPRISE apps (Cloud Platform providers, developers, companies that want to use COMPRISE apps, etc). Below are listed some exploitation models for COMPRISE:

- An instance of the COMPRISE Cloud Platform could be exploited by one or more COMPRISE partners (or a new company created for this purpose) to provide the cloud services (i.e., dataset storage and model training) and a repository of trained models that different companies can use (e.g., app developers, app providers, other voice technology companies).
- The COMPRISE Cloud Platform could be operated by third-party voice technology companies that will set up their own instance and manage both the data collected by their own voice apps and the data collected by the apps created by other companies (e.g., app developers that are generating skills for the voice technology company's generic voice assistant).
- The COMPRISE Cloud Platform could also be exploited by companies seeking app providers to develop apps for them, and that are privacy-oriented and robust enough to acquire the skills required to operate the Platform themselves.

The role of each of the parties involved in the data processing operations and their influence over the means and purposes of such operations should be assessed depending on the exploitation model employed. The parties involved may have their own purposes (e.g., developers, cloud providers, or voice assistant companies may process the same dataset for different purposes that they have decided or just process the data on behalf of another party following its instructions).

7 Privacy by design and by default

7.1 Privacy by design

The concept of "privacy by design" was introduced in Deliverable D5.1 (see Section 4.5.1) as a prevention model that demands a proactive attitude from the controller and

must be considered from the very beginning (i.e., design phase) when a service, an application or a product involves the processing of personal data.

This section delves into the concept of "privacy by design", considering the most recent guidelines published by the EDPB and different supervisory authorities. It tries to identify how the application of privacy by design may affect the design of voice-enabled technologies and the future exploitation of the COMPRISE solution.

The implementation of privacy by design requires the organisations' commitment to processing personal data and should be an integral and inseparable part of their systems, applications, products, services, and business practices and processes. Consequently, organisations should[34]:

- Promote the implementation of the privacy by design principle from the highest levels of the Administration.
- Develop a culture of commitment and continual improvement amongst the workers.
- Assign concrete responsibilities to the different members of the organisation.
- Develop methods to detect bad practices by using indicators.

COMPRISE
In accordance with the previous section, privacy by design requires the implementation of technical measures to ensure privacy (such would be the case of the COMPRISE speech and text transformation tools) and other organisational measures. Below are presented some examples of organisational measures that could be implemented during the COMPRISE exploitation stage: • Offer training on privacy and ethical use of data to the COMPRISE users. • Sign confidentiality and ethical agreements with individuals or organisations with access to the anonymised datasets (e.g., commitment not to try to re-identify data subjects). • Implement access policies.

To effectively implement data protection principles, the data controller needs to integrate adequate safeguards to ensure their efficacy throughout the life cycle of the personal data processed.[35] In the case of technologies being developed, it would be necessary to detect tactics/strategies to be followed in the different stages of the personal data processing lifecycle to make them "private by design" and ensure privacy. Engineers can employ such tactics to bridge GDPR principles and the implementation of privacy in concrete solutions[36].

To properly implement the privacy by design principle, engineers are invited to follow the steps below:

1. Specify the privacy functionalities the system should fulfil.

[34] Agencia Española de Protección de Datos. (2019). "A Guide to Privacy by Design". Retrieved from: https://www.aepd.es/sites/default/files/2019-12/guia-privacidad-desde-diseno_en.pdf
[35] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".
[36] European Data Protection Board. (2020). "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default". Retrieved from: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en

2. Design the architecture and implement the necessary elements to carry out privacy functionalities.
3. Verify whether privacy functionalities are working properly.
4. Verify the fulfilment of the privacy requirements and whether the stakeholder's expectations are being met.

Regarding step 1, please note that the standard for determining the system's privacy functionalities is the GDPR. Organisations need to implement adequate functionalities that enable compliance with GDPR principles. Organisations themselves should decide the measures to be applied, so they also manage to decide the functionalities that will be more effective to comply with GDPR principles. Of course, some private consultancy companies offer their own certification systems as a private service.

Voice assistant manufacturers and voice-enabled app developers should implement privacy by design and by default principles and integrate adequate safeguards to ensure compliance with GDPR principles during the whole lifecycle of personal data processing. As voice technologies rely strongly on deep learning and require massive processing of personal data to improve, the integration of the following safeguards should be considered both at the software and hardware level:

- Wake-up word to initiate the interaction
- On-device processing
- Encryption
- Anonymisation
- Sharing restrictions
- Privacy options activated by default in the configuration
- Guest mode option

The latest EDPB guidelines on "privacy by design" stress that the measures and safeguards adopted to implement data protection principles in personal data processing operations should be effective.[37] Data controllers must be able to demonstrate such effectiveness over time.

COMPRISE
The possibility of demonstrating the effectiveness of measures/safeguards to implement data protection principles in personal data processing operations is something the data controllers using the COMPRISE ecosystem (COMPRISE Text Transformer and COMPRISE Voice Transformer) should consider. Said effectiveness should be demonstrated over time to make adequate adjustments or developments to improve them if necessary. Therefore, for each implemented safeguard/measure, it is recommended to[38]: • Ensure that the safeguard is designed to be robust and that it is possible to scale it up when needed if the risk of non-compliance with GDPR principles increases.

[37] European Data Protection Board. (2021). "Guidelines 02/2021 on Virtual Voice Assistants, Version 1.0".
[38] European Data Protection Board. (2020). "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default". Retrieved from: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en

- Be able to demonstrate that the safeguard has achieved the desired effect. For this purpose, specific KPIs can be set and monitored.

COMPRISE has carried out several experiments (see WP2 deliverables) to test the effectiveness of its anonymisation tools. However, it is important to keep this effectiveness along time. Once COMPRISE solution enters the exploitation stage, it would be highly recommended to track effectiveness, for instance, by setting privacy-related KPIs and control the fulfilment of these KPIs, providing new versions of the tools when needed (to upgrade their effectiveness), or new safeguards or additional tools.

Also, developers should implement their own additional safeguards that need to be monitored and updated, or even integrate new safeguards when necessary.

In accordance with Article 25 of the GDPR, when implementing the necessary technical and organisational measures to comply with GDPR principles, the data controller should take into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks associated to it. Each of these elements is briefly analysed below[39]:

- The state of the art should be continuously assessed as it is a dynamic concept.
- The cost of implementation should consider both the economic cost and the efforts dedicated (i.e., in terms of human resources).
- The scope, context and purpose of the processing refer to the size and range of the processing and the data subject's expectations, depending on the purposes of the processing. In the case of voice technologies, expectations regarding privacy safeguards are high as these technologies are employed in very private environments (e.g., the data subject's home) and have the potential to collect massive amounts of data, including sensitive data (e.g., data revealed in interactions with the voice assistant, other recordings such as background conversations or sounds that may reveal different aspects of the individual's private life).
- A risk assessment should be performed in all the stages of the personal data processing life cycle. For this purpose, it is recommended to carry out a Privacy Impact Assessment (PIA), though, in certain scenarios, it is mandatory. You can find more information on how to perform a PIA in Deliverable D5.1 (Section 5).

COMPRISE

In the case of COMPRISE, existing anonymisation technologies and the possibility of improving and updating the privacy-driven tools according to the possibilities offered by the state of the art should be periodically assessed.

On the other hand, as one of the main values offered by COMPRISE is privacy, the users' expectations on the anonymisation effectiveness may be very high, so it would be convenient to inform them about the risks of not achieving total anonymisation.

Information security is another aspect that should be considered and ensured for successfully implementing privacy by design principle. When designing a product, a service,

[39] European Data Protection Board. (2020). "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default". Retrieved from: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en

or an organisation's internal process involving the processing of personal data, the confidentiality, integrity, and availability of the personal data processed should be considered. For this purpose, all the different stages of the personal data processing lifecycle should be thoroughly analysed to integrate adequate measures in each of them to ensure information security. See Section 9.2 for more information on Security.

7.2 Privacy by default

The concept of "privacy by default" is directly related to that of "data minimisation". Article 25 of the GDPR requires data protection by default, meaning that only personal data necessary for specific purposes previously defined should be processed.[40] This implies that the default settings of any application, system, product or service must be established by default to the level that offers higher protection to user's privacy.

COMPRISE
In the case of COMPRISE, the ideal situation would be that both developers and users are able to set different levels of anonymisation at some point. In accordance with GDPR's privacy by default principle, the highest level of anonymisation should be established by default. Ideally, the user of the voice app would be able to choose different levels of privacy (e.g., deciding on the categories of personal data that they would like to anonymise). However, he/she should receive information on how each level of privacy impacts the app's utility.

7.3 Anonymisation as a privacy by design technique

Deliverable D5.1 analysed the concept of "anonymisation" and explained how it differs from "pseudonymisation". It also established that the project's approach to anonymisation (anonymisation of text and speech is one of the main values brought by COMPRISE) should be aligned to that of the "Ethics and Data Protection" document of the European Commission: "*As far as your research proposal is concerned, if there is a significant prospect of re-identification of persons whose data have been collected, the information should be treated as personal data [….]*". Consequently, assessing the re-identification risk would be necessary.

During the period between the submission of Deliverable D5.1 and the preparation of this document, a thorough analysis of the anonymisation requirements and re-identification risks has been carried out. Listed below are some of the conclusions extracted from this analysis:

- Supervisory authorities diverge on the interpretation of Recital 26 of the GDPR regarding the level of risk considered tolerable after anonymisation to consider datasets as non-personal data. For example, the WGA29[41] seems to consider that

[40] European Data Protection Board. (2020). "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default". Retrieved from: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en

[41] Article 29 Data Protection Working Party. (2014). "Opinion 05/2014 on Anonymization Techniques". Retrieved from: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

no remaining risk of re-identification is tolerable for data to be considered anonymised while other supervisory authorities like the Irish Data Protection Commissioner (DPC)[42] consider it enough to demonstrate that re-identification is highly unlikely given the specific circumstances.

- The risk of re-identification will never cease to exist, as achieving a total, irreversible anonymisation is virtually impossible. In this sense, whenever anonymisation is applied to a dataset, performing a re-identification risk assessment becomes critical. Though there are different approaches to performing this assessment, some more demanding than others, accepting a tolerable risk of re-identification after the anonymisation technique is applied is most realistic.
- The assessment results should serve as a compass to decide whether additional measures should be implemented to safeguard the anonymised dataset, depending in particular on whether the dataset should be treated as personal or non-personal data, with all the implications in terms of compliance that the former implies.

The re-identification risk assessment considers several elements, such as possible identifiers and quasi-identifiers contained in the dataset, possible means of re-identification, sources where public information may be gathered, recorded information and previous knowledge that may lead to the re-identification of the data subject, the data subject profile, the environment, potential intruders and their motivations, etc. In the end, all of them will serve the data controller to decide on the optimal ways to safeguard, share or disclose the anonymised dataset, as well as to pinpoint weaknesses that should be corrected to improve anonymisation. The goal is to prevent the anonymised dataset from revealing the data subject's identity when combined with external information, for instance through linkage.

COMPRISE
Reducing the risk of re-identification depends to a great extent on the data controller, which has to analyse the datasets to decide, amongst other tasks, which data should be removed. However, the scenario has proven to be quite different for COMPRISE, where the anonymisation of the datasets is automatic, hindering the re-identification risk assessment. COMPRISE must focus on expanding the number of identifiers and quasi-identifiers it could remove or substitute for addressing this issue. Additionally, the possibility of an intermediary step (human intervention) could be considered to assess re-identification risks before making the anonymised datasets public and sharing them in the Cloud Platform.

8 International transfers of data

Section 4.6 of Deliverable D5.1 explains international transfers of personal data to third countries as well as the requirements that need to be fulfilled to comply with the GDPR.

Between the publication of Deliverable D5.1 and the preparation of the present document, significant changes directly impacting the requirements for international transfers

[42] Data Protection Commission. (2019). "Guidance on Anonymization and Pseudonymisation". Retrieved from: https://www.dataprotection.ie/sites/default/files/uploads/2020-09/190614%20Anonymization%20and%20Pseudonymisation.pdf

of personal data have taken place. These changes are expected to affect several companies transferring personal data to third countries (outside of the EEA) on a regular basis (e.g., companies storing data in servers located outside of the EEA).

As voice technologies typically operate as cloud-based services (the user's speech is sent to the cloud, where it is automatically transcribed and processed, and the system's reply is sent back to the user's device), voice technology companies may be storing the personal data collected in the servers of storage providers located in third countries.

Another scenario involving an international transfer of personal data could be that of a voice app/skill developer who acts as a data controller to personal data collected through the app and a voice assistant designer who acts as a data processor (see Section 5.1 of this document). When users interact with the assistant, their voice goes through the servers of the voice assistant's designer to be transcribed as text and interpreted, which might be located in a third country.

In this regard, the Court of Justice of the European Union (CJEU) adopted the Judgment (C-311/18) of 16th of July 2020, better known as the Schrems II decision[43], which invalidated Decision 2016/1250 on the adequacy of the protection provided by the EU-US Data Protection Shield.

Additionally, in the judgement, the CJEU sets out a burden on data exporters who wish to use Standard Contract Clauses approved by the EC (see Section 4.6 of Deliverable D5.1). The data exporter must carry out an assessment taking into account the circumstances of the transfer to verify whether that level of protection is respected in the third country concerned (in practice, a level of protection that is essentially equivalent to the one guaranteed by the GDPR).

The EDPB has published some recommendations to warrant an appropriate level of privacy in international transfers of data for public consultation.[44] According to these recommendations, each case should be analysed individually by the data controller, and the following steps should be taken:

1. Map all transfers of personal data to third countries (i.e., be aware of where the data goes).
2. Verify the tool the transfer relies upon, amongst those listed under Chapter V of the GDPR (an adequacy decision, safeguards listed in Article 46 or derogations provided in article 49).
3. Analyse the level of protection offered by the third country, focusing on its data protection legislation to find any elements that may influence the safeguards' effectiveness.
4. Adopt the measures necessary to bring the level of protection of the data transferred up to the EU standard of essential equivalence (i.e., implement supplementary measures to the safeguards of Article 46). The complementary measures should be chosen by assessing the specific context of the international transfer.

[43] Judgment of the Court (Grand Chamber) of 16 July 2020. "Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems". Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62018CJ0311

[44] European Data Protection Board. (2020) "Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data". Retrieved from: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/recommendations-012020-measures-supplement_en

5. Take any formal procedural steps the adoption of supplementary measures may require.
6. Periodically re-evaluate the level of protection afforded to the data transferred to third countries.

Supervisory authorities will consider the actions taken by the exporters to ensure a protection level similar to the one offered by the GDPR.

9 Information security and cybersecurity

As voice-based systems continue to conquer new niches, cybercriminals have been perfecting their techniques and adapting them to target this type of technology specifically.

The ubiquity of voice-based systems has given the companies behind them access to extremely valuable data (e.g., medical data, financial data) cybercriminals can employ for a variety of malicious purposes, including ransomware attacks or impersonation.

This section aims to describe both cybersecurity trends that can be implemented to secure data in voice-enabled devices and cybersecurity threats that may pose a risk for personal data and sensitive information recorded by voice-enabled devices. Besides, it aims to extend and update the information provided in Section 6 of Deliverable D5.1.

9.1 General cybersecurity threats

Over the years, cyberattacks have become more sophisticated to match cybersecurity advances. However, this doesn't mean that traditional attacks cannot affect state-of-the-art technologies in any way.

Before analysing cybersecurity threats specific to voice technologies, an overview of some of the most used techniques employed by cybercriminals in 2019-2020 according to Crowdstrike 2020 Global Trend Report[45] is provided below:

- **Scripting**: Type of attack where malicious scripts are injected into otherwise secure and trusted websites[46].
- **Disabling security tools**: The configuration of security tools is modified or directly disabled to prevent them from running, allowing attackers to access the system undetected.
- **System Owner/ System Discovery**: This attack aims to identify the primary user or groups of users who use a system or whether they are actively using a system to target them.
- **Account Discovery**: Type of attack where adversaries attempt to get a listing of accounts of a system in a determined environment.
- **Registry Run Keys / Start folder**: Adversaries achieve persistence by adding a program to a start-up folder or referencing it with a registry run key.

[45] Crowdstrike. (2020). “Crowdstrike 2020 Global Trend Report”. Retrieved from: https://www.crowdstrike.com/resources/reports/2020-crowdstrike-global-threat-report/
[46] Splunk. (2020) “Top 50 Security Threads”. Retrieved from: https://www.splunk.com/pdfs/ebooks/top-50-security-threats.pdf

- **Process Injection:** Adversaries inject code into a process to evade process-based defences, allowing access to the process's memory, system/network resources or elevated privileges[47].
- **Hidden Files and Directories**: Adversaries utilise files and directories that don't show up unless explicitly requested for persistence, evade system analysis, etc.
- **PowerShell**: Adversaries utilise PowerShell as an entry point, gaining persistence and elevated privileges.
- **Credential Dumping**: Aims at obtaining account login and password information, typically in the form of a hash or a clear text password, perform lateral movement and access restricted information[48].
- **Masquerading**: Adversaries borrow legitimate user and system identities to trick victims into submitting valuable information (e.g., personal data).

Furthermore, according to McAfee's Cloud Adoption and Risk Report, there has been a 630% rise in cyberattacks on cloud services since January 2020. The health industry is the second most affected right behind financial services[49].

The State of the Cloud 2020 from security specialist Sophos found that, in 2019, 70% of organisations hosting data or workload in the public cloud experienced a security incident, with multi-cloud organisations reporting up to twice as many incidents in comparison to single platform adopters. As for the type of attacks, the same research found that around 34% of organisations were hit by malware of some nature, 29% experienced exposed data, 28% were target to ransomware attacks, 25% found their accounts compromised, and 17% were subject to crypto-jacking[50].

It is also important to point out that, in the particular case of voice-based systems, most of the personal data recorded by the user device is kept in the cloud, which, unfortunately, offers attacker multiple remote ways of accessing it (e.g., web/app-enabled access). Besides, cloud environments are heavily exposed to insider attacks (e.g., abuse of authorised access) and suffer from incomplete data deletion, enabling the system to retain private information after intending to delete it[51].

The attacks described above are only a sample of the repertory cybercriminals employ for malicious purposes. Though most of them cannot directly affect voice-based devices such as smart speakers, they can target their users or the companies that manage services offered by the device.

For instance, smart speakers require individuals to provide an email address to create a user account. In a masquerading attack, an attacker can send phishing emails to the user tricking them into believing they are from a legitimate source (e.g., the developer of

[47] Mitre. "System Owner/User Discovery". Retrieved from: https://attack.mitre.org/techniques/T1033/
[48] Red Canary. (2020). "2020 Threat Detection Report". Retrieved from: https://redcanary.com/threat-detection-report/techniques/credential-dumping/
[49] McAfee. (2020). "Cloud Adoption and Risk Report". Retrieved from: https://www.mcafee.com/enterprise/en-us/assets/reports/restricted/rp-carr-wfh.pdf
[50] Sophos. (2020). "The State of the Cloud Security 2020". Retrieved from: https://secure2.sophos.com/en-us/medialibrary/Gated-Assets/white-papers/sophos-the-state-of-cloud-security-2020-wp.pdf
[51] Edu J., Such J., Suarez-Tangil G. (2020). "Smart Home Personal Assistants: A Security and Privacy Review". Retrieved from: https://arxiv.org/pdf/1903.05593.pdf

an Alexa Skill) and request sensitive information such as credit card numbers or contact details.

In this regard, in 2019, a fake Amazon Alexa skill tricked enough users to reach No. 60 on the Top Free apps section of the Apple iOS's AppStore, and No. 6 in the Top Ten lists for Utilities. The fraudulent skill, called "Setup for Amazon Alexa", and developed by a company called One World Software, asked users to provide, among other information, their IP addresses, which can reveal user's city, state or province and ZIP code, and the serial number of the Alexa device they intend to set up, all highly valuable for cybercriminals[52].

In the same way, an attacker could modify or completely disable the configuration of the security tools of a service provider platform to achieve persistence and access sensitive information available (e.g., about users) without being detected, for instance, by injecting codes into poorly secured subdomains (e.g., track.amazon.com, used to track packages), making lasers to pass as voice commands, and other techniques.

Lastly, sometimes cybercriminals don't require sophisticated techniques to access a system. Undetected vulnerabilities (e.g., subdomain vulnerabilities) can facilitate attackers work as it occurred to Amazon voice-based devices last year. In 2020, a flaw in Amazon Alexa home devices allowed hackers to access personal information (i.e., user profile, including home address) and user conversation history through a malicious Amazon link specifically created for this purpose. Once the user clicks the link, the attacker could get a list of all installed Alexa skills and a token that allowed them to download or delete skills at discretion. The attackers used the token to remove legitimate skills and replace them with malicious ones that employ the same invocation phrase. This way, they could easily access personal/sensitive data or use the user's online services at their expense[53].

Now, as previously mentioned, there are cyber threats that specifically target voice-based systems. Before detailing some of the most important, below are listed some recommendations both to reduce the risks of being subject to cyberattacks and the impacts of cyberattacks[54]:

- Given that approximately 40% of security breaches are indirect, i.e., threat actors target weak links in the business ecosystem and supply chain, organisations must consider extending cybersecurity measures to the ecosystems surrounding their enterprises.
- Organisations must invest in resources to identify and fix breaches faster. Cyber recovery/restoration time and response time should also become a priority to organisations to reduce the impact of cyber threats.
- Organisations must learn to identify the most suitable solutions and technologies to deal with cyber threats, e.g., SOAR (Security, Orchestration, Automation, Re-

[52] Iribarren M. (2019). "Fake Alexa Setup App in Apple Store Removed After Climbing Charts". Retrieved from: https://voicebot.ai/2019/01/03/fake-alexa-setup-app-in-apple-store-removed-after-climbing-charts/

[53] BBC News. (2020). "Amazon Alexa security bug allowed access to voice history". Retrieved from: https://www.bbc.com/news/technology-53770778

[54] Accenture. (2020). "Innovate for Cyber Resilience: Lessons from Leaders to Master Cybersecurity Execution". Retrieved from: https://www.accenture.com/_acnmedia/PDF-116/Accenture-Cybersecurity-Report-2020.pdf

sponse) technologies are preferred by organisations to achieve faster incident responses and shorter recovery times, while for faster incident detection, AI is preferred.

- Organisations must collaborate with strategic partners to share knowledge of threats and test their cybersecurity resilience.
- Organisations must collaborate in the creation of cybersecurity policies and standards and implement security awareness training.
- Periodically perform advanced testing, including vulnerability assessments and routine penetration testing. The creation of data loss prevention (DLP) programmes is also encouraged.

9.2 Cybersecurity threats that affect voice-based system

As the popularity of voice systems grows, so do illegitimate attempts to access the data collected by them. Deliverable D5.1 briefly addressed voice squatting, software attacks, and audio adversarial attacks. This section aims to describe other threats that target voice-based systems, with largely known security and privacy issues (i.e., weak authentication, weak authorisation, adversarial AI, underlying and integrated technologies, traffic analysis — particularly when multiple smart appliances are connected to the voice-based system, etc.) are exploited by attackers around the world.

Most of the attacks described in this section target the user device (e.g., a smart speaker, a smartphone, etc.). However, attacks could target other elements, including the[55]:

- **Voice-enabled system service provider cloud**: The attack targets the voice-enabled system cloud components, exploiting ML vulnerabilities and underlying technologies.
- **Third-party web skills/actions/apps:** Attackers exploit user misconceptions about the voice-enabled system, and in particular, about the skill/action/app.

9.2.1 Surfing Attack

SurfingAttack is a type of inaudible attack that utilises ultrasonic guided waves to elicit a reaction from smartphone voice assistants. It works over solid mediums/materials and allows for multi-round interactions with the device, given that voice assistants in smartphones require users to input a command (e.g., asking a question) before performing the action (e.g., answering the question)[56].

SurfingAttack allows attackers, among others, to perform fraudulent calls using the victim's smartphone, interact with connected devices using the voice assistant, and retrieve the victim's SMS verification codes[57].

[55] Edu J., Such J., Suarez-Tangil G. (2020). "Smart Home Personal Assistants: A Security and Privacy Review". Retrieved from: https://arxiv.org/pdf/1903.05593.pdf
[56] Owaida A. (2020). "Voice assistants can be hacked with ultrasonic waves". Retrieved from: https://www.welivesecurity.com/2020/03/04/voice-assistants-hacked-ultrasonic-waves/
[57] Yan Q., Liu K., Zhou Q., Guo H., Zhang N. (2020). "SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves". Retrieved from: https://surfingattack.github.io/papers/NDSS-surfingattack.pdf

It is particularly dangerous for users who secure their accounts with two-factor authentication and use SMSs to receive their authentication codes, as attackers could get access to the user's online services and, thus, their sensitive data or private information.

SurfingAttack have proven successful in various devices of major brands, including Apple, Huawei, Xiaomi, Google and Motorola.

9.2.2 Laser attacks

Lasers can be used to hijack voice assistants in some smartphones and smart speakers (Apple HomePod, Amazon Echo, Apple iPhones, Google Home) by vibrating sensors in the micro-electro-mechanical system (MEMS) microphones in the device.

In these attacks, the device's microphones interpret the light as a voice command, making it vulnerable to the attacker's malicious intentions. However, to be successful, the attacker is required to have a direct view of the smartphone or speaker, and the laser must be aimed at a very specific part of the microphone.

It is important to note that this type of attack is partly possible because voice-controlled systems rarely require users to authenticate themselves (e.g., with passwords or PINS).

Laser attacks could allow the attacker, among others, to unlock smart lock-protected doors, locate, unlock and start smart vehicles that are connected, for instance, to the victim's Google account, and shop on online stores at the victim's expense.

9.2.3 Long-range attack

Both SurfingAttack and DolphinAttack (see Section 6.5 of Deliverable D5.1) are inaudible ultrasound attacks that only work from close ranges, primarily because of a property of acoustic hardware (microphone and speakers) called non-linearity, which causes high signal frequencies to shift to lower signal frequencies.

To send a high-frequency signal to a smart speaker (or any other targeted device), the sound must be played through a speaker. Non-linearity will make voice commands transmitted over inaudible ultrasound frequencies shift into lower audible bands after passing through the non-linear acoustic hardware.

DolphinAttack is not affected by non-linearity because it works at low power, which limits its range to approximately 5 ft[58].

In this sense, a group of researchers from the University of Illinois has developed a new speaker design (LipRead) that facilitates long-range attacks by splicing the spectrum of the voice command into segments and playing each of them on different speakers limiting the leakage from the speaker. In short, by portioning the spectrum, the leakage would be below the threshold of human hearing, thereby preventing users from hearing the attacker's fraudulent commands[59]. Once refined, this system could serve attackers to target devices that are out of their sight range.

[58] Colyer A. (2018). "Inaudible voice commands: the long-range attack and defense". Retrieved from: https://blog.acolyer.org/2018/05/11/inaudible-voice-commands-the-long-range-attack-and-defense/

[59] Roy N., Shen S., Hassanieh H., Choudhury R. "Inaudible Voice Commands: The Long-Range Attack and Defense". Retrieved from: https://synrg.csl.illinois.edu/papers/lipread_nsdi18.pdf

9.2.4 Network attacks

Network attacks are regularly performed by a Man-In-The-Middle (MITM) that cautiously analyses network traffic to inject commands for malicious purposes. In voice-enabled systems, these attacks might occur during the speaker setup with the ASR/TTS server to steal user information like passwords or other access permissions[60].

9.3 Deepfakes

Artificial intelligence, machine learning and neural networks, together with the increasing capacity of computer systems, have made it possible for cybercriminals to mimic voices in what is known as a "deepfake". According to the CyberCube report "Blurring reality and fake: A guide for the insurance professional", criminals focus on two main areas when it comes to deepfakes[61]:

- **Voice conversion**. A technique that involves the sampling of two voices, a source and a target, and the application of software to convert one into another, and
- **Text to speech**. A technique that allows a mimicked, synthesised voice to be instructed to say whatever the user of the software commands via a text interface[62].

Cybercriminals can employ both techniques for diverse purposes. For instance, to obtain personal information from a third party (e.g., credit card numbers, telephone numbers, addresses, etc) by mimicking someone's voice.

9.4 Approaches to address cybersecurity threats associated with voice technologies

The previous subsections analysed different types of cyberthreats, including more sophisticated attacks that specifically target voice technologies. This section presents a series of countermeasures and recommendations that users, developers and even manufacturers can implement to reduce the risks associated with them.

9.4.1 How can users address cyber threats?

Some attacks that target voice-based systems and devices can be prevented or mitigated by users by taking simple precautions. For instance, users can defend against SurfingAttack by turning off screen personal results (on Android) and disabling their voice assistant on the lock screen. Even by using thicker phone cases made of uncommon materials like wood[63].

---

[60] Park Y., Choi H., Cho S., Kim Y., (2019). "Security Analysis of Smart Speaker: Security Attacks and Mitigation". Retrieved from: http://www.techscience.com/cmc/v61n3/35289/pdf

[61] CyberCube. (2020). "Blurring reality and fake: A guide for the insurance professional". Retrieved from : https://insights.cybcube.com/hubfs/Reports/Social-Engineering_CyberCube_report.pdf

[62] CyberCube. (2020). "Blurring reality and fake: A guide for the insurance professional". Retrieved from : https://insights.cybcube.com/hubfs/Reports/Social-Engineering_CyberCube_report.pdf

[63] Yan Q., Liu K., Zhou Q., Guo H., Zhang N. (2020). "SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves". Retrieved from: https://surfingattack.github.io/papers/NDSS-surfingattack.pdf

Below are listed a series of measures users can put in practice to reduce the risks of being targeted by cybercriminals[64]:

- Add an additional layer of authentication (e.g., a voice PIN) to voice-based devices.
- Disable mics to prevent data from being collected without consent — for instance, Amazon Alexa offers an alternative called Alexa Voice Remote that allows the user to press and hold the remote's talk button and then input the command.
- Keep sound notifications on, so that the user is alerted if the device is accidentally triggered.
- Periodically delete all interactions with the voice-based system.
- Keep the device away from windows and doors, where cybercriminals can easily access it through SurfingAttack, laser attack, etc.
- Double-check the origin of apps/skills/actions before downloading them, as they can serve as a point of access to cybercriminals.
- Avoid entering subdomains, which could also serve as a point of access for cyber-criminals.
- Always read third-party terms and conditions before starting to use an app, skill, action, etc.

9.4.2 Implementation of a Software Development Life Cycle (SDLC)

Despite the precautions mentioned above, one of the best countermeasures against cyber threats (and, overall, to ensure security) is to consider them from the first moment a device or system is conceived, that is, from design. This approach shifts the responsibility to developers who are, most of the time, more versed in cybersecurity than the average user. The Software Development Life Cycle (SDLC) has been developed with that purpose in mind.

Software is the most critical part of every IoT system, including voice-enabled systems. It enables their functionalities and provides value added features. However, given its nature, it is exposed to multiple security risks that may compromise the system's overall security and the services it provides.

The SDLC is a multiple-phase process that aims at delivering effective and efficient services based on the design and the requirements of the system. It allows improving the system's overall security by considering security across all phases of the SDLC, from the beginning of the software development process to its maintenance and posterior disposal, and by implementing appropriate security measures where necessary.

Below are presented some consideration regarding the different phases of the SDLC[65]:

- **Requirements**: In this phase, the user, business, legal, regulatory and functional requirements (i.e., physical/hardware requirements for the development) of the software are defined. Business use cases must be considered in the requirements phase and target environment specifics, and other context-related matters.

[64] Safety Team. (2020). "10 Quick Tips for Amazon Echo Safety & Privacy". Retrieved from: https://www.safety.com/amazon-echo-safety/

[65] European Union Agency for Cybersecurity. (2019, November). "Good Practices for Security of IoT". Retrieved from: https://www.enisa.europa.eu/publications/good-practices-for-security-of-iot-1

- **Software design**: In this phase, a set of documents describing how the user/business and functional requirements will be implemented in the system's specifications must be drafted. The security models established in the requirements phase are reviewed (and refined if needed) using threat modelling.
- **Development and implementation**: In this phase, the previously established requirements and designs are transposed into a programming language. The code should be built, tested and integrated (and subsequently maintained and updated) with security in mind. This could be achieved by implementing continuous security best practices and assessments.
- **Testing and acceptance**: This phase assesses if the resulting product (e.g., a voice-based device) met all the requirements established in the requirement phase. Identifying the specific needs of the device or system to establish the most suitable testing strategy is crucial in this phase.
- **Deployment and integration**: In this phase, all the necessary elements of the software in the production ecosystem and its deployment are integrated. Choosing an adequate strategy for the deployment strategy is crucial, especially if multiple actors are involved.
- **Maintenance and disposal**. In this phase, the strategy to maintain the availability and integrity of the product and its functionalities must be established. Continuous vulnerability assessments and penetration tests are necessary to ensure security and adequate threat detection and repose. All these measures should be maintained even if maintenance functions are delegated to third parties. Lastly, erasure mechanisms must be provided to ensure compliant disposal.

Voice assistants would benefit from an SDLC. Not only would it allow developers and engineers to assure that the requirements set at the beginning are the project met expectations or whether they need to be refined, but also to define how security will be handled once the project ends, particularly the processing of personal data in the platform.

The SDLC would also help establish which third parties will have access to personal data, how they will handle this data, and whether legal requirements are being properly addressed.

9.4.3 Technical cybersecurity measures for voice-based systems

Lastly, we present a series of technical cybersecurity measures that are intended to mitigate common risks associated with voice-based systems proposed in the research article "Smart Home Personal Assistants: A Security and Privacy Review"[66]:

- **Voice authentication**: Voice authentication can help reduce the risks associated with voice-based systems' weak authentication. This measure helps the system tell apart individual users when they speak.
- **Location verification**: Location verification can also help reduce risks associated with weak authentication. It allows voice-based systems to verify whether a user is truly nearby before executing voice commands.
- **Frequency Filtering and Spectral Analysis**: Aiming at enhancing authentication, this measure seeks to protect voice-based systems against synthesised speech by employing frequency filtering and spectral analysis techniques.

[66] Edu J., Such J., Suarez-Tangil G. (2020). "Smart Home Personal Assistants: A Security and Privacy Review". Retrieved from: https://arxiv.org/pdf/1903.05593.pdf

- **Traffic shaping**: This measure aims to defend voice-based systems against profiling. It works by routing the traffic between different gateways routers of multiple cooperating smart appliances before sending it to the internet.
- **Jam the Speakers**: To limit the amount of data shared with companies behind voice-based devices (sometimes unintentionally), but also to prevent attackers from input unaidable commands for malicious purposes, gadgets generate noise and interference close to smart speakers' microphones can be used to jam them. The user must establish a wake-up word to indicate to the gadget when to stop the jamming to allow the speaker to listen and respond to voice commands[67].
- **Background filtering technologies**: To remove background noise that may contain personal information such as third parties (different from the voice assistant user), conversations (background conversations) should be considered to implement technologies filtering the unnecessary data and ensuring that only the user's voice is recorded.

Please note that no cybersecurity measure is infallible: these measures are only meant to reduce risk, not to eliminate it completely.

10 Ethics

This section aims to extend and update the Ethics sections of Deliverable D5.1. For that purpose, it provides helpful insights on ethics concerns related to the use of artificial intelligence and, more specifically, of voice technologies, including those related to the design of voice-based systems; actual and future capabilities of AI-based technologies; initiatives carried out in the European Union to address ethics in AI, and efforts to legislate on this matter.

10.1 Ethical concerns related to AI and voice technologies

Smart speakers (e.g., Google Nest, Amazon Echo Dot, etc.) and smartphone voice assistants (e.g., Siri) are perhaps the most common voice-based systems used by the average user.

For instance, in 2019, the favourite activities of American smart speaker users included asking the device questions, listening to streaming music services, and checking the weather forecast. Other popular use cases included setting alarms or timers, listening to the radio, using an Alexa Skill or Google action, playing games, controlling smart home devices, listening to news or sports, etc[68].

Of course, many other activities can be performed with intelligent voice assistants, such as checking bank accounts, checking the traffic, accessing calendars, accessing apps, etc. What they all have in common is that they could reveal personal data and sensitive information about the user and third parties could be present when they input a command. What companies do with this data is a source of concern.

[67] Paranoid. "Paranoid Home Devices". Retrieved from: https://paranoid.com/products
[68] Statista. (2019) "Smart speaker use case frequency in the United States as of January 2019". Retrieved from: https://www.statista.com/statistics/994696/united-states-smart-speaker-use-case-frequency/

Moreover, the poor understanding the average user has on how a voice-based system works (e.g., what data they record, how its behaviour is audited, to what extent their responses are explainable, etc.)[69] exacerbates this issue.

Listed in the following subsections are some scenarios related to the use of voice-based technologies that carry ethical implications.

10.1.1 Data Ownership

Who owns the voice data recorded by the device? How can this data be used? For instance, in 2015, authorities demanded the release of all recordings from an Amazon Echo device to investigate the murder case of an American citizen called Victor Collins[70].

10.1.2 Societal biases

Data used to train machine learning applications could learn societal biases. For instance, if a user types "O bir hemşire" or "Or bir doctor" into Google Translate, the system will translate it into "She is a nurse" and "He is a doctor" despite "o" being a gender-neutral pronoun in Turkish. This bias is believed to originate from the presumption that doctors are always males and nurses females, which, in turn, reflects on the training data on which Google Translate is built.

On the other hand, societal biases can take many forms, racial biases being among the most common. In this regard, the study "Biased bots: Human prejudices sneak into AI systems" showed that in typical training data used for machine learning, African American names are often used alongside unpleasant words (e.g., "hatred", "poverty", "ugly"), while European American named, on the contrary, are used more often with words such as "love", "lucky" or "happy"[71].

Moreover, even the way the user speaks could trigger biases in a voice-enabled system. For instance, AI can infer that a strong accent correlates with a poorer education, leading the system to provide these groups of users (i.e., non-native speakers) with simpler or dumber responses[72]. It is also possible for AI to learn to recognise speech with certain accents better than others. This type of bias is known as the "accent gap" and regularly affect non-native speakers or regional speakers (i.e., native speakers with strong accents characteristic of a specific region, poorly represented in the datasets used for training the system).

Even the fact that voice assistants regularly have female voices could result from developers being mostly male. However, some studies suggest that individuals, in general,

[69] National Pilot Committee for Digital Ethics. (2020). "Ethical Issues on Conversational Agents". Retrieved from: https://www.ccne-ethique.fr/sites/default/files/cnpen-chatbots-call-participation-2020-11-10.pdf

[70] Jarosciak J. (2017). "Social and Ethical Concerns of Smart Voice-Enabled Wireless Speakers". Retrieved from: https://www.joe0.com/2017/03/29/social-and-ethical-concerns-of-smart-voice-enabled-wireless-speakers/

[71] University of Bath. (2017). "Biased bots: Human prejudices sneak into AI systems". Retrieved from: https://www.bath.ac.uk/announcements/biased-bots-human-prejudices-sneak-into-ai-systems/

[72] Cox. T. (2019). "The Ethics of Smart Devices That Analyse How We Speak". Retrieved from: https://hbr.org/2019/05/the-ethics-of-smart-devices-that-analyse-how-we-speak

prefer female voices over male voices, which is why many companies opt for the feminisation of voice-based technologies[73]. According to a study by Indiana University that played male and female synthesised voices for men and women, both groups reported female voices sounded 'warmer'[74]. Another study from Stanford University found out that people prefer male voices when used to teach them about computers. In contrast, female voices are preferred, for example, for providing relationship advice[75].

In the end, the best solution for societal biases is to diversify the datasets used for training AI models. Today, many crowdsourcing initiatives aim to improve voice technologies, for which hundreds of hours of recordings from users of all genders and from all around the world are made available to developers (e.g., Mozilla Common Voice).

10.1.3 Anthropomorphisation of voice-enabled systems

The intention of tech companies to provide assistants with human-like personalities leads users to anthropomorphise voice-enabled systems and, thus, overshare sensitive information, including personal data. This phenomenon, which ascribes human attributes to machines, increases the risks of deteriorating human self-determination and leads users to overestimate and inflate the system's capabilities.

Developers seek conversational agents to gain user trust, which is the primary reason some systems can tell jokes, sing songs, or even mimic emotions like sadness, for instance, if the user said something disrespectful or mean.

Solutions proposed to address this issue include raising awareness of the machine nature of smart speakers and similar devices. Instead of focusing on the system's human-like characteristics (e.g., the possibility of holding a conversation or asking it trivial questions), the user should be strongly informed of the actual functions and advantages the device has[76].

10.1.4 Niche in sensitive fields

AI has found a niche in the field of personal health, rich in personal data and sensitive information (e.g., history of the patient medical diagnoses, diseases or interventions, medications prescribed, test results, behavioural patterns, sexual life, etc.).

For instance, the emergence of software and ICT tools for medical practitioners allow for the development of decision support systems, which improve the individual capacities of these professionals. In this sense, we will have voice-enabled systems specially designed to record medical appointments or domestic voice-based smart devices (e.g., smart speakers) that remind users when to take medications, all of which can be used by medical practitioners to provide better diagnoses, monitor the patient's improvements,

---

[73] UNESCO, EQUALS Skills Coalition. (2019). "I'd blush if I could: closing gender divides in digital skills through education". Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000367416.pdf

[74] IUPUI, School of Informatics and Computing News Release (2017). "MacDorman explores voice preferences for personal digital assistants". Retrieved from: https://soic.iupui.edu/news/macdorman-voice-preferences-pda/

[75] Liberatore S. (2017). "Why AI assistants are usually women: Researchers find both sexes find them warmer and more understanding". Retrieved from: https://www.dailymail.co.uk/sciencetech/article-4258122/Experts-reveal-voice-assistants-female-voices.html

[76] Jesionowski K. (2019). "Analyzing ethical challenges of digital advertising for the Amazon Echo voice assistant". Retrieved from: http://kxjournal.com/wp-content/uploads/2019/05/Voice-Assistants-Ethics.pdf

etc. What is relevant about using ICT in the medical field is that decision-making has become a spatially distributed process, where multiple actors such as medical specialists, nurses, pharmacologists, etc., converge[77].

Although this sharing of medical data appears to have many advantages for medical practitioners (even administratively) and patients, it raises concerns regarding privacy and ethics.

Privacy concerns include the pervasiveness of ICT technologies, lack of transparency of healthcare professionals' work, and difficulty of respecting privacy and confidentiality from third parties with a strong interest in getting access to recorded data. On the other hand, ethical concerns revolve around fundamental rights, such as human dignity, which serves as a basis for requirements of privacy, confidentiality and medical secrecy; beneficence and non-maleficence, which serve as a basis for the attempts to weigh anticipated benefits against foreseeable risks, and solidarity, which serve as a basis of the right for everyone to the protection of healthcare, particularly in regards to vulnerable groups[78].

In the end, as it occurs with all of the scenarios listed, it is the uncertainty of what can be done with medical data that raise the above mentioned ethical concerns. However, it is possible to implement some measures to reduce privacy-related risks. Below are listed a few examples:

- The use of encryption techniques of sensitive data.
- The use of pseudonymisation or, if possible, anonymisation techniques on personal data.
- On-device processing (to the extent possible) or, if not possible, avoidance of public clouds.
- Strong agreements with third parties that may access personal data (e.g., from patients).

10.1.5 Absence of human intervention

Algorithms are sometimes used for delicate tasks such as determining how much an individual should pay for insurance or filter candidates applying for a position. Although these tasks can be performed more efficiently with the help of AI, they require the strict supervision of human beings, which is not always fulfilled.

For instance, several speech recognition solutions in the market claim to successfully identify fraudulent call centre conversations and even criminals pretending to be customers. Without enough human intervention to verify that they are indeed criminals or scammers, these systems could end up wrongly labelling legitimate users as such.

[77] European Commission. (2019). "Ethics Guidelines for Trustworthy AI". Retrieved from: https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf
[78] European Commission. (2019). "Ethics Guidelines for Trustworthy AI". Retrieved from: https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

10.1.6 CNPEN ethical factors

The French National Pilot Committee for Digital Ethics (CNPEN), on the other hand, listed a series of ethical factors related to the use of voice-based systems that are briefly described below[79]:

- **Status confusion**: Relates to the anthropomorphisation of voice-based systems. This ethical issue revolves around the possibility for users to confuse conversational agents with human beings, given that they often mimic human traits.
- **Naming**: Developers consider that addressing conversational agents by a name improves their functioning. In some fields (e.g., healthcare), getting to name voice-based systems is positive for the user, heightening their emotional responses. However, some ethical issues arise from this possibility. Who should name the device, the designer or the user? Is it necessary to assign it a gender? Is it correct to assign it a human name? Would a non-human name be better?
- **Bullying of voice-enabled systems**: Humans tend to project their qualities onto voice-based systems. In this sense, it is not uncommon for humans to insult or mistreat conversational agents. Most of the time, voice-based systems are programmed to respond according to strategies set by the designer. Several ethical considerations arise from this scenario. Is insulting voice-based systems a morally reprehensible act? Should voice-based systems be able to respond by insulting the user in turn? If a voice-based system is assigned a feminine identity, should abuse be addressed as abuse towards women?
- **Trust in voice-enabled systems**: Developers seek to establish and maintain the user's trust in the system as it is necessary for the device to perform its functional tasks. However, humans must be aware that an excess of trust in the system could be detrimental to them. One of the main ethical concerns arising from this particular scenario is whether conversational agents should introduce themselves as the user's assistant, advisor or friend.
- **Conflicts with voice-enabled systems**: Conflicts may arise from the interaction between the user and the conversational agent. For instance, a user that deeply trusts a conversational agent may feel troubled after receiving an answer it considers incorrect or a lie. In this sense, are lies from a conversational agent more or less acceptable than a human lie? What are the limits to this capability?
- **Manipulation by voice-enabled systems** – **Nudge theory**: Based on the American Nobel Prize in Economics winner Richard Thaler, the concept of nudge involves encouraging individuals to change their behaviour without coercing them by using cognitive biases. In the case of voice-based devices, this is achieved through suggestions or manipulations designed to influence the user's behaviour or emotions (e.g., a voice assistant can be used as a mean of influencing individuals for commercial or political purposes). Ethical concerns around this factor are clear. Are all nudges allowed? Is it possible to distinguish between good and bad nudges? Does nudge invalidate the concept of free and informed consent?
- **Free choices**: This ethical issue particularly reflects on recommendation systems, where a single choice (made by the conversational agent after evaluating all possible answers to the user's command) could limit the user capacity to choose freely by depriving it of accessing the full range of available options. In this regard, ethical

[79] National Pilot Committee for Digital Ethics. (2020). "Ethical Issues on Conversational Agents". Retrieved from: https://www.ccne-ethique.fr/sites/default/files/cnpen-chatbots-call-participation-2020-11-10.pdf

issues revolve around the possibility of allowing the conversational agent to explain all or several choices to the user.

- **Emotions of the system**: Concerns revolve around the possibility of building conversational agents capable of detecting human emotions and/or that simulate human emotions.
- **Vulnerable individuals**: Voice-based systems may occupy vulnerable individuals' (e.g., kids or adults with intellectual disabilities) full attention as it replaces standard human socialisation. However, to what extent is this acceptable and for what purposes? Could this type of interactions lead to a lasting change in vulnerable individuals' lifestyle or social interactions?
- **The memory of the dead**: Post-mortem use of conversational agents often occurs, though the right to privacy is supposed to end when the individual dies. When is it acceptable for voice-based systems to reproduce the voice of a deceased individual?
- **Surveillance**: Voice-enabled systems capable of recording voices can monitor users' interactions around them, whether with humans or other conversational systems. This functionality poses severe ethical and legal risks related to the protection of privacy, the use of personal data without consent, professional secrecy, etc. In this sense, it is logical to wonder under which circumstances the disclosure by conversational agents of content recorded is permitted?
- **Work**: There are several ethical concerns regarding the use of conversational agents in working environments, including how they can change the evolution of certain professions, by what means should the use of conversational agents be regulated or simply if the use of conversational systems should be encouraged or prohibited for certain professions or human practices.
- **Long-term effects on language**: The use of voice-enabled systems could have a lasting impact on human language and lifestyle habits. In this case, ethical concerns revolve around whether the influence of conversational agents should be deemed positive or negative.

Lastly, the CNPEN has also reflected on ethical issues related to the design of voice-based systems, including specification problems, metrics and evaluation functions, goals assigned to the system, training biases and instability, explainability and transparency, and the impossibility of rigorous evaluation due to the black box phenomenon[80].

10.2 Present and future capabilities of AI in voice-based systems

Today, artificial intelligence is present in a multiplicity of technologies, as is speech recognition. As a result, on 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence.

However, there is still a lot to legislate on this matter. Artificial intelligence is a constantly evolving field. Tech giants such as Google, Amazon, Microsoft, etc. are already employing AI to enhance the capabilities of their voice-based devices, such as:

- Personalisation to provide the user what he/she wants before asking for it (e.g., the voice-based device suggests the user his/her regular payment method).

[80] National Pilot Committee for Digital Ethics. (2020). "Ethical Issues on Conversational Agents". Retrieved from: https://www.ccne-ethique.fr/sites/default/files/cnpen-chatbots-call-participation-2020-11-10.pdf

- Contextualisation (e.g., to provide the user with an answer coherent with the circumstances).
- Recognition of symptoms (e.g., to aid health workers diagnosing patients).
- Recognition of emotions (e.g., to seek help from third parties in case of life-threatening situations).

Large amounts of sensitive information can be collected from the user in all four scenarios presented above, though most of them are still far from perfect. In this sense, can users expect to receive personalised ads based on the emotions detected by the voice-enabled system? Would a health worker be allowed to share medical data with third parties (i.e., pharmaceutical companies) based on the symptoms detected by the system?

Although none of the aforementioned situations are specifically regulated, several working groups and organisations have devoted themselves to creating guides and recommendations for the use of artificial intelligence in its multiple fields of application, including that of voice technologies, as seen in the upcoming subsections.

10.3 Regulation of ethical issues in the EU

The European Commission has made it clear that the use of AI and technologies in general must be aligned with the fundamental rights set out in international human rights laws, the EU Treaties and the EU Charter as expressed in the Ethics Guideline for Trustworthy AI.

According to the these guidelines, for AI to be considered trustworthy, it must comply with the following principles[81]:

- **Human action and supervision**: AI systems should empower humans, enabling them to make informed decisions, while ensuring at the same time proper oversight mechanisms, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches. For instance, if a voice-based system is used as a means to diagnose a user, whatever diagnostic it makes should be reviewed by a human specialist before any action is taken.
- **Technical security and robustness**: AI systems should be resilient and secure to ensure the prevention or minimisation of unintentional harm. For instance, voice assistants must be equipped with measures to act against fraudulent access to the user's personal data or information related to their interactions or prevent them (e.g., double authentication mechanisms).
- **Data and privacy management**: Privacy, data protection, and adequate data management should be ensured, considering the quality and integrity of the data. For instance, many voice assistants now allow users to delete their history of commands and decide whether or not their voice recordings are used to train and improve the system. Both Deliverable D5.1 and the present document provide recommendations related to privacy in voice-based systems.
- **Transparency**: AI systems and their decisions should be explainable to the users concerned (knowing their capabilities and limitations). For example, if a voice-based device is employed for diagnosing COVID-19, users should be aware of the

[81] European Commission. (2019). "Ethics Guidelines for Trustworthy AI". Retrieved from: https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

factors the system considers before providing a diagnosis and the overall logic behind the model (if not a black box). Moreover, users should also be aware of the details concerning the data being collected by the device.

- **Diversity and non-discrimination**: AI systems should be accessible to all and avoid biases. For instance, voice-based devices should be trained using diverse datasets to avoid gender biases (better understanding of male voices over female voices) and racial biases (poor understanding of users from certain ethnic origins).
- **Social and environmental wellbeing**: AI systems should benefit all human beings, be sustainable and environmentally friendly, and consider their social impact. For instance, voice technologies could be employed for educational purposes or help the disabled (e.g., a voice-based device that helps the blind use public transportation, have access to books and other educational material, etc.) amongst other uses that bring a very positive impact on society.
- **Accountability**: The responsibility and accountability of AI systems and their outcomes should be ensured through adequate mechanisms.

In the case of voice technologies, these principles can be fulfilled, for instance, by training STT and NLU algorithms in a privacy-preserving way to provide improved interaction functionalities and enable communication in several languages and dialects, instead of providing decisions that may affect the user.

Moreover, these principles must be assessed throughout the entire lifecycle of the AI-based system, given that even legitimate processing activities could result in unexpected outcomes (i.e., collateral impacts that may affect data subjects).

10.4 Other initiatives related to ethics

This last subsection shortly explores various initiatives (at European level) that aim to address ethics in the use of various technologies, especially Artificial Intelligence (AI):

- **The European Observatory on Society and Artificial Intelligence** (OSAI) is a project created under the Horizon 2020 Programme that aims to offer tools to help people better understand and study the impact of AI technologies have across the EU. Although it does not intend to enact laws or guidelines on ethics and artificial intelligence, the observatory supports the discussion of AI's ethical, legal, social, economic and cultural issues within Europe[82].
- **"The ethics of artificial intelligence: Issues and initiatives"**: In March 2020, the European Parliament published a study on the ethical implications and moral questions that arise from the development and implementation of AI technologies. The study also delves into the different guidelines and frameworks that countries inside and outside the European Union have created to address the use of AI technologies and compares the current frameworks and main ethical issues related to AI[83].

[82] The AI4EU Observatory. Retrieved from: https://www.ai4eu.eu/ai4eu-observatory
[83] European Parliament. (2020). "The ethics of artificial intelligence: Issues and initiatives". Retrieved from: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf

11 Conclusion

This document provided an in-depth analysis of good practices to successfully implement GDPR principles, extending the work done in Deliverable D5.1. It explored recent guidelines published by Supervisory Authorities, research work and good practices adopted by voice technology companies.

The GDPR principles and their application to voice technologies are enriched with use cases. In this sense, some scenarios provided by the CNIL and the EDPB where personal data processing operations through voice technologies and the legal bases for each of them are analysed. The document also reviewed scenarios where voice-enabled systems record individuals without their consent or any other lawful legal basis, a common scenario these days.

Readers will find helpful the list of the categories of data that voice assistants and voice apps may collect, the information that could be extracted from each of them, and the recommendations provided on how to inform the data subject on this matter properly. Common purposes for processing personal data through voice assistants or voice apps were included as well.

The importance of training voice-enabled systems with diverse datasets to avoid biases and comply with the fairness principle and good practices implemented by voice technologies companies to minimise personal data processing through voice systems (e.g., wake word detection, guest modes, etc.) are also explored.

Readers would also encounter some functionalities that voice assistants provide to facilitate the exercise of data subject's rights, such as deleting their recordings or revoking permissions to third parties. The identification of the data controller and data processor in voice-enabled systems where parties usually fulfil more than one role is also analysed through multiple use cases presented by the CNIL.

The document also deepens on the concepts of privacy by design and by default, adopting a more practical approach than its predecessor. It presented recommendations for engineers on how to comply with privacy by design and anonymisation as a privacy by design technique. Regarding the latter, it was stressed the importance of performing a re-identification risk assessment to determine whether the risk of re-identifying the data subject is tolerable or non-tolerable and, consequently, decide how data should be treated (i.e., as personal data or non-personal data). The particular scenario of COMPRISE, where anonymisation is automatic, was also studied. A series of recommendations to decrease the risk of re-identification (e.g., expanding the number of identifiers and quasi-identifiers the solution could remove or substitute) is presented.

The document also analysed changes in the international transfer of data caused by the elimination of the EU-US Data Protection Shield, which affects voice technology companies operating in Europe, as well a list of cybersecurity threats, including attacks specifically designed to affect voice technologies, and how users, manufacturers and developers could reduce the risks if being affected by them.

It presented mechanisms and recommendation to comply with the GDPR principles when personal data is processed through voice technologies that are applicable to COMPRISE, more specifically, to its exploitation phase. It also delved into ethical concerns around voice-enabled systems, which employs technologies such as Artificial Intelligence and Machine Learning that, improperly used, could severely affect the data subject's privacy (e.g., oversharing personal information). Issues such as the ownership of

the data collected through voice-based systems or the existence of biases leading to discriminatory treatments and European initiatives to address ethics are studied as well.

In conclusion, several practices voice technology companies engaged with aren't compliant with the GDPR. However, over the past years, good practices and technical and organisational solutions to improve this flaw have emerged. This document intends to provide the best mechanisms to comply with the GDPR and protect data subjects.

Appendix A. Privacy preserving options in apps

The AEPD provides some examples of the operations that could be included in the privacy panel configured by the users. As indicated before, the most privacy-preserving options should be activated by default, and the user should be able to change this configuration if it considers it appropriate[84]. Below are listed some of these measures:

- Operation in anonymous mode.
- Operation without the need to create a user account.
- Operation with different user accounts on the same device for the same data subject.
- Operation with different user accounts on different devices for the same data subject and data processing (e.g., collection, storage, manipulation sharing or any other operation over personal data.).
- Identification by means of tools and technologies that enhance privacy, such as attribute-based credentials, zero-knowledge tests, etc.
- Alternatives and wilfulness in the contact information requested to the user: e-mail, postal, telephone.
- Monitoring techniques in processing (cookies, pixel tags, fingerprint, etc).
- Configuration of unique identifiers (tracking IDs), programming of their reset and notification of activation times.
- Device metadata collected from the device (battery consumption, OS, versions, languages, etc).
- Metadata included in the processed or generated media (in documents, photos, videos, etc).
- Information collected about the user's internet connection (device from which they connect, IP address, data from the device sensors, application used, browsing and search log, web-page request, date and time log, etc.) and information regarding elements located near the device (Wi-Fi access points, mobile phone service antennas, activated Bluetooth devices, etc).
- Information collected about user activity on the device: powering up, activating applications, using the keyboard or mouse, etc.
- Free activation and deactivation of data collection systems (cameras, microphones, GPS, Bluetooth, Wi-Fi, movement, etc).
- Establishing a temporary schedule of when sensors (e.g., cameras, microphones, etc.) can be operational.
- Physical blockers (such as camera lens cover tabs, speaker blockers, etc).
- Exercise of the right to object, right of imitation or right to erasure.

[84] Agencia Española de Protección de Datos. (2020). "Guidelines for Data Protection by Default". Retrieved from: https://www.aepd.es/sites/default/files/2020-10/guia-proteccion-datos-por-defecto-en.pdf

- Possibility of configuring all optional processing operations for non-essential purposes, such as data processing for service improvement, usage analysis, ad customisation, detection of usage patterns, etc.
- Incorporation of a user data reset option to resume the relationship from scratch
- Specific section for configuration options related to sensitive data.
- Help and transparency panel with examples of use and possible risks and consequences for the rights and freedoms of the user.
- Incorporation of a specific means (button or link) to return to the initial configuration with default values.
- Configuration of session data erasure after closing.
- Configuration of maximum time limits for logging out of the application or devices.
- Time limits for the storage of user profiles.
- Control of the erasure of temporary copies.
- Elimination of the user's trace in the service: "Right to be forgotten".
- "Right to be forgotten" mechanism regarding the information published in social networks or other systems.
- Configuration of historical data storage periods in the service: e.g., on purchase sites, last articles, last consultations, etc.
- Profile information of the data subject shown to the user and third parties: name, pseudonym, telephone number, etc.
- Information about the status of the data subject accessible to third parties. For example, in messaging applications, information on availability, message writing, message reception, message reading, etc.
- Automatic session blocks.
- Control of data storage encryption.
- Control of data communication encryption.
- Control of print output deletion.
- Alerts regarding the connectivity status of devices.
- Configuration of notices and reminders to data subjects about what information dissemination and disclosure policies are in place.
- Control the distribution scope of the information distributed in the application environment (social networks, work networks, etc).
- Configuration of the reception of warnings when information is being made accessible to third parties.
- Control of the metadata incorporated in the information generated or distributed.
- Choice of where personal data is stored, either on local or remote devices and, in the latter case, other parameters such as processors or countries.
- History of profiles and entities that have accessed your information.
- Information on access to your data by authorised users.
- Information on the latest changes carried out and the profile that has made the change.

Appendix B. COMPRISE's feedback to the EDPB guidelines on Virtual Voice Assistants

The H2020 COMPRISE project welcomes the opportunity to provide input to the European Data Protection Board's (EDPB) consultation on its draft guidelines on virtual assistants ("Guidelines 02/2021 on Virtual Voice Assistants") published on 9th of March 2021.

COMPRISE is an H2020 project financed by the European Commission that defines a fully private-by-design methodology and tools that reduce the cost and increase the inclusiveness of voice interaction technology through research advances on privacy-driven data transformations, personalised learning, automatic labelling, and integrated translation. This leads to a holistic easy-to-use software development kit interoperating with a cloud-based resource platform. The sustainability of this new ecosystem is demonstrated for three sectors with high commercial impact: smart consumer apps, e-commerce, and e-health.

Below, please find our comments on the guidelines:

Section	Paragraph	Text	Comment
Executive summary		"Currently, all VVAs require at least one user to register in the service. Following the obligation of data protection by design and by default, VVA providers/designers and developers should consider the necessity of having a registered user for each of their functionalities."	Please, consider including an example of one or two functionalities for which it wouldn't be necessary for the user to register. It would help to make the paragraph clearer.
Section 2.2	16	"Please note that while currently most voice-related processing is performed in remote servers, some VVA providers are developing systems that could perform part of this processing locally".	Please, consider including in the footnote examples of open source European initiatives such as COMPRISE, which also perform part of the processing locally (on device or a personal server).
Section 2.5	21	"The over or under-representation of certain statistical characteristics can influence the development of machine learning-based tasks and subsequently reflect it in its calculations, and thus in its way of functioning, just as much as its quantity, the quality of the data plays a major role in the finesse and accuracy of the learning process."	The consequences of the under-representation of certain population segments in the training datasets can be illustrated with an example. One clear consequence of underrepresenting population segments that particularly affects voice assistant users is "the accent gap", i.e., the inability of voice-based technologies to understand speakers with non-native or regional accents with the same accuracy as most speakers. Also, consider analysing the bias issue in voice technologies and compliance with the "fairness principle". A subsection could be added to Section 3.
Section 3.1	30	“If data controllers become aware (e.g., through automated or human review) that the VVA service has	Please, consider including a recommendation stating that data controllers should main-

		accidentally processed personal data, they should verify that there is a valid legal basis for each purpose of processing of such data. Otherwise, the accidentally collected data should be deleted."	tain a proactive attitude regarding the performance of reviews to identify possible accidental recordings of personal data.
Section 3.1	31	"Moreover, it should be noted that personal data processed by VVAs may be highly sensitive in nature. It may carry personal data in its content (meaning of the spoken text) and meta-information (sex or age of the speaker etc.). The EDPB recalls that voice data is inherently biometric personal data. As a result, when such data is processed for the purpose of uniquely identifying a natural person or is inherently or determined to be special category personal data, the processing must have a valid legal basis in Article 6 and be accompanied by a derogation from Article 9 GDPR (see Section 3.8 below)."	Please consider indicating that very sensitive information may be inferred through the user's voice and the existence of patented technologies that aim to infer the user's health status and emotional state from its voice.
Section 3.2	36	"The plurality of personal data processed when using a VVA also refers to a plurality of personal data categories for which attention should be paid (see below Section 3.8). The EDPB recalls that, when special categories of data are processed, Article 9 GDPR requires the controller to identify a valid exemption from the prohibition to processing in Article 9(1) and a valid legal basis under Article 6(1), using an appropriate means identified under Article 9(2). Explicit consent may be one of the appropriate derogations where consent is the legal basis relied on under Article 6(1)."	There may be voice apps that, at first glance, do not seem to request any sensitive data directly or for which the purpose of the processing does not require the collection of sensitive data. However, they still collect sensitive information or allow for sensitive information to be inferred (e.g. a cooking app through which the user asks for specific ingredients that may reveal their health condition or an e-commerce voice app through which the user may acquire products that may reveal their health status, sexual orientation, or religious beliefs). Apps that enable very open interactions could lead to the user revealing sensitive information (e.g., a voice app for writing a diary, a voice app to

			write notes or input appointments in the calendar). For these cases, consider providing some guidelines on the best way to proceed for the data controller (e.g., inform the user about the possibility of collecting these kinds of sensitive data and asking for explicit consent, dataset anonymisation, etc.)
Section 3.3	48	"Failure to provide necessary information is a breach of obligations that may affect the legitimacy of the data processing. Complying with the transparency requirement is an imperative since it serves as a control mechanism over the data processing and allows users to exercise their rights. Informing users properly on how their personal data is being used makes it more difficult for data controllers to misuse the VVA for purposes that go far beyond user expectations. For example, patented technologies aim to infer health status and emotional states from a user's voice and adapt the services provided accordingly."	When humans carry out labelling, the need for transparency could be included in this Section as an example. There is a general perception that voice technology companies have failed to inform their clients adequately on the processing of their personal data. Several media published in 2019 hinted that different voice technology companies failed in informing their clients that they were hiring humans to review clips of conversations between devices and their users.
Section 3.2.2	58	"VVA designers must consider how to properly inform non-registered and accidental users when their personal data is processed. When consent is the legal basis for processing users' data, users must be properly informed for the consent to be valid. In order to comply with the GDPR, data controllers should find a way to inform not only registered users, but also non-registered users and accidental VVA users. These users should be informed at the earliest time possible and at the latest, at the time of the	Is there any good practice or mechanism for informing non-registered users and accidental VVA users of personal data processing by a VVAA that could be provided as an example?

		processing. This condition could be especially difficult to fulfil in practice".	
Section 3.6	96	"The data minimisation principle is closely related to the data storage limitation principle. Not only do data controllers need to limit the data storage period, but also the type and quantity of data."	Please, consider including some guidelines to determine the criteria that the data controller should follow to decide how long personal data should be stored when this data is processed through voice technologies.
Section 3.6	105	"Anonymising voice recordings is especially challenging, as it is possible to identify user through the content of the message itself and the characteristics of voice itself. Nevertheless, some research is being conducted on techniques that could allow to remove situational information like background noises and anonymise the voice".	The two articles cited in the footnote are irrelevant. The paper by Cohen-Hadria et al. does not "remove situational information like background noises". On the contrary, it aims to preserve background noise and obfuscate any overlapping speech. The method by Qian et al. provides almost no protection, as we showed recently.[85] Please consider citing the voice anonymisation baseline for the 1st VoicePrivacy Challenge[86] or the open-source voice [87] and text[88] anonymisation tools developed by COMPRISE as example tools that provide much more effective anonymisation.
Section 3.6	107	"Before considering anonymisation as means for fulfilling the data storage limitation principle, VVA providers and developers should check the anonymisation process renders the voice unidentifiable."	Please consider citing COMPRISE's rigorous evaluation protocol[89] (based on formal informed attacker models combined with state-of-the-art voice biometrics) as an example solution to check whether the voice is unidentifiable. Also, clarify that effective anonymisation decreases the utility of the data (i.e., its suitability for training ASR or NLU models), although this impact is limited for

[85] Srivastava B. M. L., Vauquier N., Sahidullah M., Bellet A., Tommasi M., Vincent E. (2020). "Evaluating voice conversion-based privacy protection against informed attackers", in 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2802-2806.
[86] Voice Privacy Challenge. "Introducing VoicePrivacy". Retrieved from: https://www.voiceprivacychallenge.org/
[87] COMPRISE. "Voice Transformer". Retrieved from :: https://gitlab.inria.fr/comprise/voice_transformation
[88] COMPRISE. "Text Transformer". Retrieved from: https://gitlab.inria.fr/comprise/text_transformer
[89] COMPRISE. (2021). "Deliverable D2.3 "Final transformation library and privacy guarantees". Retrieved from: https://www.compriseh2020.eu/files/2021/02/D2.3.pdf

			some anonymisation techniques.
Section 3.9	140	"VVA designers should consider technologies deleting the background noise and conversations ensuring that only the user voice is recorded."	The article cited in the footnote is irrelevant. It does not delete the background noise nor background conversations. On top of that, it provides almost no protection against re-identification, as explained above. Deleting the background noise or background conversations requires using speech enhancement technology (with special attention to privacy), which has not been done so far to the best of our knowledge.
Footnotes	• Footnote 5 • Footnote 34 • Footnote 47		The URL link is broken (due to line break). The URL address in the link after the line break is missing.