WHITE PAPER COLLECTION N°1

ON THE RECORD

Exploring the ethical, technical and legal issues of voice assistants

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

WHITE PAPER COLLECTION **- N°1**

ON THE RECORD
Exploring the ethical, technical and legal issues of voice assistants

CONTENTS

November 2020

Publication Director: Gwendal Le Grand

Editor-in-chief: Bertrand Pailhès

Editors of this white paper: Martin Biéri and Félicien Vallet with the help of Isabelle Corbara, Pauline Faget, Basile Guley, Estelle Hary and Juliette Hirtz.

Graphic design: Agence Linéal - +33 (0)3 20 41 40 76

Printing: DILA

Legal Deposit: for publication

Illustrations : Malte Müller

The views expressed in this publication do not necessarily reflect the position of the CNIL.

The following are warmly thanked Monir Azraoui, Nacéra Bekhat, Régis Chatellier, Clément Commerçon, Antoine Courmont, Eric Delisle, Hugo Dussert, Louis Dutheillet de Lamothe, Pauline Girard, Jérôme Gorin, Armand Heslot, Sonia M'Jati, Mathias Moulin, Aymeric Pontvianne, Flora Sanchez and Clémence Scottez as well as ARCEP, CNPEN, CSA and Hadopi.

EDITORIAL

Even though they have been a must-have in recent years, voice assistants are opening up a new era in the use of the digital tools that surround us, from our telephone to our television set, our vehicle or our household appliances. An era that would be conversational and interactive and that could revolutionise the way we communicate, consume, access information... and therefore live.

By enabling more seamless interactions, simplifying orders and offering ease of use, such devices could bring about real progress in digital and social inclusion. They could become embedded in users' day-to-day routines, particularly people in a dependent situation, whether elderly or disabled.

However, these undeniable advances should not obscure the questions that voice assistants raise from a data protection perspective, in particular from the point of view of transparency in the way their system functions.

First of all, the raw material they handle is very strongly anchored in our private lives. As the first tool for communicating with our fellow human beings, the voice is in fact an integral part of our identity and reveals, in addition to the meaning of words, a great deal of information about the speaker. Age, gender, physical condition, accent, geographical and socio-cultural origin, education, health or emotional state, but also identity – voice being a biometric characteristic allowing identification – are all examples of information that can be deduced from the audio signal. Current events have illustrated the risks linked to accidental eavesdropping on individuals' private spaces and attest to the indispensable trustworthiness that we must be able to expect from the manufacturers of these objects, but also from those who deploy them and those who contribute content for them, by creating applications for example.

Second, while there is a widespread adage in the business that voice assistants can enable "technology to disappear", the use made of data should never be made invisible.

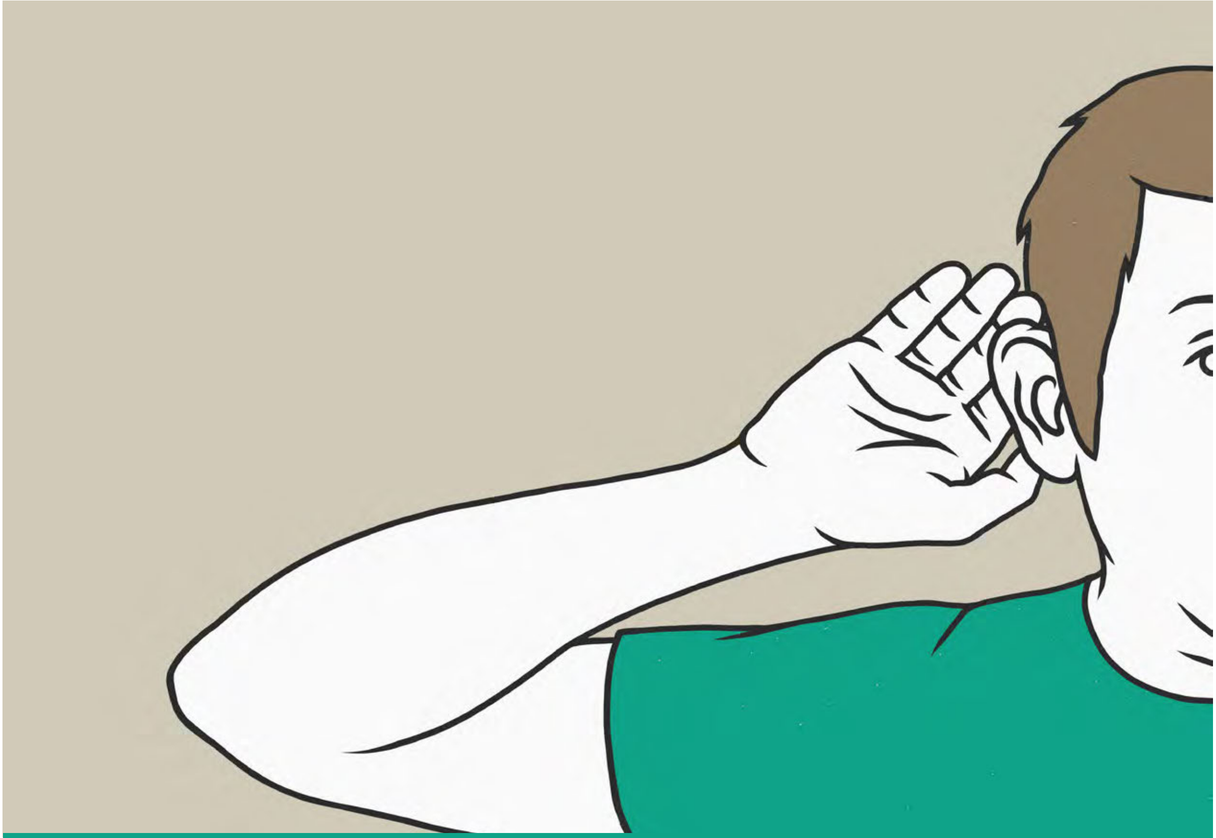
Beneath their practical and entertaining aspect is the reality that some of these objects capture our lifestyle habits to enhance a profile, which can end up locking us into a commercial ecosystem. Do these new advantages outweigh their disadvantages? If so, how can they be regulated to ensure they respect the privacy of individuals?

The CNIL has been exploring these questions for several years and from different angles. Since 2017 and the arrival in our homes of voice assistants built into living room smart speakers, the CNIL has published numerous articles and interviews with experts on its institutional website (www.cnil.fr) or on that of LINC, the CNIL's Digital Innovation Laboratory (linc.cnil.fr), dedicated to forward-looking analyses, studies and experiments. The CNIL has monitored the development of these devices by being in contact with the various designers and, through numerous partnerships such as the one with Inria (the French National Institute for Research in Digital Science and Technology), supports the research works being carried out on the subject.

The aim of this white paper is thus to make this work accessible to all types of audience. The aim is to present the various legal, technical or ethical issues, and to respond to the concerns of those who build these assistants, those who deploy them, as well as those who use them. Finally, it also aims to offer advice and guidance to help ensure that these tools develop in a way that respects the fundamental rights of their users.

It is hoped that this white paper will provide some avenues for reflection on the use of data by voice assistants, in particular compliance with the main principles advocated by the GDPR, and contribute to a collective thought process as required by this new relationship with digital technology.

Marie-Laure Denis, *President of the CNIL*



WHAT'S THE STORY BEHIND VOICE ASSISTANTS?

While it has been possible to record sound – and therefore voice – for almost a century and a half, speech is still associated with a certain volatility. However, the widespread use of automatic speech processing technologies and their integration into a growing number of objects is creating a new relationship with the "voice object". This all seems to point to an essential paradigm shift for users.

WHAT'S SO SPECIAL WITH THE VOICE?

**Speak, exchange, communicate, say, enunciate, tell, converse, etc.
There are many terms to characterise the oral exchanges we have on a daily basis.
However, our voice is still a great unknown.**

A private piece of data

The voice is a key building block as far as our identity is concerned. It is a concrete expression of language, which is itself a communication skill. It is in particular this complex spoken language ability that distinguishes us from other animal species. In practice, the voice conveys, apart from speech (the words themselves), many of the speaker's characteristics: emotions, intentions, physical condition, etc. By relying on perception mechanisms, our listeners are able to interpret these signals and decipher these states.

A distinction is generally made in communication between "high level" activities, integrating intellectual processes (perception, conception of speech, decoding and analysis of the message, etc.) and "low level" activities, allowing the exchange to actually take place (physical mechanisms of articulation and sound production, hearing, etc.). We can therefore consider the voice as a "low-level" process since it is a physical tool at the service of thought and discourse. In 1916, Ferdinand de Saussure introduced the notion of a linguistic sign which illustrates this state of affairs[1]. It unites "not a thing and a name, but a concept and a sound-image", the sound-image being called signifier and the concept it designates signified.

Thus, our voice carries more complex messages than the simple "meaning" of the words spoken (the signified). This non-verbal information contains the so-called paralinguistic elements: silences, intonations, gestures, body posture, tone of voice, facial expressions, etc. The voice signal thus allows the extraction of many different types of information: the meaning of the message, of course, but also indications relating to age, gender, physical condition, familiarity with the language, accent, geographical and socio-cultural origin, education, state of health or emotional state, etc.

A question to...
Joana Revis

Throughout our lives, our voice accompanies us. What relationship(s) do we have with it?

Our voice is so much a part of us that it's the very first thing we do at birth: cry out! It really does accompany us throughout our lives, it changes over time, it evolves, it is there all the time without us thinking about it and most of us end up having a rather ungrateful relationship with it: it's just there and we take it for granted – we don't really look after it. In the end, there are only two situations in which we become aware of it: when we have a close-knit relationship with it (this is for example the case for singers or actors), or when we lose it (during simple laryngitis or in the case of chronic damage to the vocal cords). Then, all of a sudden, we realise how important it is.

Joana Revis is a speech and language therapist-vocologist and associate lecturer at Aix Marseille University's Faculty of Medicine

> Full interview to be found on LINC [in French]
Joana Revis, *Notre voix porte en elle toutes les intentions qui sont les nôtres*, Linc.cnil.fr, march 2018,
https://linc.cnil.fr/fr/
joana-revis-notre-voix-porte-en-elle-toutes-les-intentions-qui-sont-les-notres

[1] - Ferdinand de Saussure, *Course in general linguistics*, 1916

Seen as a mere element of non-verbal language, the voice is often little considered for what it is. For Joana Revis, on the contrary, it occupies a central place in our communication-based society[2]. Our voice is a mirror of the soul, capable in particular of singing, laughing, demanding, enunciating, convincing, crying, reassuring, reproaching, imploring, consoling, warning, manipulating, playing and thus expressing what is unspeakable, betraying emotions, or characterising a personality. The voice is never unequivocal. It varies over time for the same individual while remaining profoundly singular. It is normal under these conditions that we develop very strong relationships with it and with the voices we hear around us.

A volatile piece of data

Another of the great specificities of the voice is its intangible yet volatile nature. Physically the voice is only a trace left by air movements caused by the phenomenon of phonation, i.e. the production of sounds specific to the spoken language. However, since the invention of the phonograph by Thomas Edison in 1877, it has been possible to record these traces on recordings and subsequently replay and analyse them. As technology has progressed, the voice has become very easy to capture – even potentially without speakers knowing it[3].

Every individual has personality rights that are intended to protect him or her. Among these rights, the right to privacy and the image right are recognised as two distinct, subjective rights intended to protect moral integrity. While the former was affirmed in 1948 by the United Nations Universal Declaration of Human Rights (Article 12) and in French law in 1970 in Article 9 of the Civil Code, it was in the 19th century, contemporaneous with the invention of sound recording (and photography) that the image rights was recognised. When dealing with image rights in general, it is common to associate them with the visual image. However, as stated by Paris Regional Court (TGI) on 19 May 1982 in the context of a lawsuit initiated by the singer Maria Callas following the unauthorised broadcasting of work recordings, "the voice is an attribute of the personality, a sort of sound image"[4]. In a similar vein to the right to a person's image, the right to a person's voice must also therefore be taken into account.

Article 226-1 of the Criminal Code provides that "any person who wilfully violates the privacy of another person's private life by capturing, recording or transmitting, without the consent of their author, words spoken in a private or confidential capacity shall be punished by one year's imprisonment and a fine of 45,000 euros". This provision bestows comprehensive protection over all words spoken in a private setting, images belonging to or depicting a person, information relating to his or her home or places s/he attends, information relating to his or her state of health, private letters and e-mails, information relating to his or her family life or love life, or his or her political, religious or philosophical opinions. Various elements of case-law exist and make it possible to characterise this infringement.

In practice, we are often interested in the combination of different criteria, namely: the clandestine nature of the recording, its location, duration, etc.

> «
> ***The voice is never unequivocal. It varies over time for the same individual while remaining profoundly singular.***
> »

[2] - Joana Revis, *La voix et soi : Ce que notre voix dit de nous*, DeBoeck, 2013.
[3] - Félicien Vallet, *Les droits de la voix (1/2) : Quelle écoute pour nos systèmes ?*, Linc.cnil.fr, May 2019, https://linc.cnil.fr/fr/les-droits-de-la-voix-12-quelle-ecoute-pour-nos-systemes
[4] - TGI Paris, 19 may 1982, aff. Maria Callas, Recueil Dalloz 1983, jurisprudence p. 183 ou encore Paris, 22 janv. 2001, Dalloz 2002, p. 2375, note A. Lepage - Adde, D. Huet-Weiller, *La protection juridique de la voix humaine*, RTD civ. 1982, p. 497

A question to... Nicolas Obin

More and more companies are offering products that allow you to create a digital clone of your own voice or that of a third party. What are the possible commercial uses?

The ability to reproduce a person's voice in a natural and realistic way has opened up many possible applications that companies have been rushing into. Computer-generated voices are poised to accompany us in our everyday routines, whether through smartphones, home assistants or in vehicles, virtual receptionists and automated call centres, to name just a few examples. Some companies now offer the possibility of creating a synthesised voice with one's own voice, at the cost of a certain number of recordings to be made. Preliminary tests were carried out as early as 2014 to be able to speak in another language but keeping one's own voice, with machine translation as a direct application. In time, it will probably be possible to create voice avatars for video game players, to create the voice of characters, or to perform automatic dubbing of an actor in different languages. Voice cloning in speech synthesis also has important medical applications, such as "voice prosthesis". All these possibilities, while fascinating and heralding the future of these technologies, are nevertheless in a largely embryonic state.

Nicolas Obin is a lecturer at the Science and Technology of Music and Sound (STMS) laboratory), Ircam - CNRS - Sorbonne University.

> Full interview to be found on LINC [in French]
Nicolas Obin, *La voix artificielle rend la machine plus humaine*, Linc.cnil.fr, march 2019, https://linc.cnil.fr/fr/nicolas-obin-la-voix-artificielle-rend-la-machine-plus-humaine

A reproducible piece of data

Another aspect of the right to a person's voice concerns "vocal identity". While the act of falsifying audiovisual content is not new, the recent possibilities offered by text-to-speech synthesis are raising questions about fraud and identity theft afresh[5]. In particular, the application of artificial intelligence technologies known as deepfake (from the association of words deep learning and fake) technologies leverage powerful machine learning techniques to manipulate or generate visual and audio content with a high potential for deception. In practice, the main methods used today are auto-encoders and generative adversarial networks (GANs)[6]. Although the current quality of voice clones produced from small quantities of audio samples does not yet seem to be satisfactory[7], the subject, which until recently was in the realm of science-fiction, is now attracting increasing attention. From a legal point of view, while cases of digital voice spoofing are still very rare[8], or unheard of, in the case of vocal imitations, the act of "misleading" listeners has been punished by the courts for almost fifty years. In 1975, a judgment sanctioned the use of a television commercial based on a text read by a person whose "diction, rate, tone and inflections of voice [...] evoked the verbal peculiarities of actor Claude Piéplu (the voice of the Shadoks)"[9]. Abroad, and particularly in the United States, similar cases were reported, for example for singers Tom Waits and Bette Midler[10]. However, imitation of a person may in certain cases be permitted, in particular if it is justified by the historical or topical context in which the work is situated, if it is not of a defamatory nature or if it is in the style of parody or caricature.

Variable geometry data

In addition to making sure in our exchanges that we understand the meaning of the message we receive, we analyse the way in which it is delivered to us. This multi-level characterisation thus reveals much richer information than a simple sequence of juxtaposed words; it is therefore, as such, a datum to be handled with care. The General Data Protection Regulation (GDPR) provides that any information relating to a natural person who can be identified, directly or indirectly, is understood to be personal data. Such a definition covers many types of data: name, registration number, telephone number, photograph, date of birth, municipality of residence, fingerprint, geolocation history, etc.

[5] - Félicien Vallet, *Les droits de la voix (2/2) - Quelle parole pour nos systèmes ?*, Linc.cnil.fr, june 2019, https://linc.cnil.fr/fr/les-droits-de-la-voix-22-quelle-parole-pour-nos-systemes

[6] - Ruben Tolosana et al., *DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection*, arXiv, january 2020, https://arxiv.org/pdf/2001.00179.pdf

[7] - Jaime Lorenzo-Trueba et al., *Can we steal your vocal identity from the Internet?: Initial investigation of cloning Obama's voice using GAN, WaveNet and low-quality found data, Odyssey,* 2018, https://arxiv.org/pdf/1803.00860.pdf

[8] - Morgan Tual, *« Deepfake » : dupée par une voix synthétique, une entreprise se fait dérober 220 000 euros*, Le Monde, september 2019, https://www.lemonde.fr/pixels/article/2019/09/06/deepfake-dupee-par-une-voix-synthetique-une-entreprise-se-fait-derober-220-000-euros_5507365_4408996.html

[9] - Claude Piéplu case, TGI Paris 3 December 1975, D. 1977, p. 211, notes R. Lindon.

[10] - Daniel Payette, *Les autres facettes de l'image: le nom, la voix et la ressemblance,* Les Cahiers de la propriété intellectuelle, 2015, https://www.lescpi.ca/s/162

In this case, information derived from the voice signal emitted by an individual may also be personal data. As a favoured vehicle of our social interactions, the analysis that we carry out of the voice can allow for the message transmitted to be understood, its sender to be identified, but also the latter to be categorised according to different modalities. It is therefore personal data which, depending on the use made of it, is of variable geometry. It must be ensured that the key principles of data protection – which concern in particular the relevance of data processing, its transparency, rights of the individuals, data control, risk management and security – are respected.

A question to...
Jean-François Bonastre

What are the future developments in voice technologies?

The voice carries a lot of information about the individual such as age, gender, origins, education, feelings, physical or mental state and perhaps even intentions... Of course, or perhaps fortunately, we do not know how to decipher all this information with sufficient reliability for proper exploitation. Not yet, at least... Investigating whether alcohol or drug use can be seen in the voice is an avenue that many labs are exploring. Finally, there are studies and possible applications revolving around the detection of emotions or emotional attitudes. Some even go so far as to assess sincerity, another name for a lie detector test... Even if scientific rigour and results are not always present, sessions on these subjects are regularly proposed during the major scientific conferences in the field, often around "challenges" pitting systems against one another and giving the impression that everything has been sorted out...

Jean-Francois Bonastre is a professor at the Computer Science Laboratory of Avignon and a specialist in speech processing and voice authentication

> Full interview to be found on LINC [in French]
Jean-François Bonastre, *La voix n'est pas une biométrie classique*, Linc.cnil.fr, february 2017, https://linc.cnil.fr/fr/jean-francois-bonastre-la-voix-nest-pas-une-biometrie-classique

One of the most obvious use of voice data is the automatic speech recognition, i.e. textual transcription of spoken words and phrases, which allows the meaning of the message to be "decoded". In practice, this involves matching the air movements picked up by a microphone with the sequence of words spoken by the person. Article 9 of the GDPR provides for a prohibition in principle to the processing of "sensitive" data, while allowing for certain exceptions, such as obtaining the consent of the data subjects. Automatic speech recognition may thus reveal some of these "sensitive" data, such as information relating to political opinions, religious or philosophical beliefs, trade union membership or sex life. Other information that may be contained in audio recordings may also be considered sensitive data and may not necessarily relate only to the words spoken. For example, various automatic processing operations may relate to racial or ethnic origin[11]. Health-related data can also be inferred. For example, research has been conducted for several years to characterise the presence of degenerative diseases such as Alzheimer's or Parkinson's[12]. Indeed, among the clinical manifestations, voice disorders are those that occur at an early stage of disease development.

Finally, sensitive data within the meaning of the GDPR also includes biometric data, i.e. data allowing or confirming the identification of an individual by his or her physical, physiological or behavioural characteristics, when its processing aims at uniquely identifying a person. They are produced by the body itself and characterise it definitively. They can sometimes be used to track and identify an individual, even without their knowledge. This data is particularly sensitive because it is permanent. What's more, it can in many cases be captured remotely, without the person's knowledge, for example for voice or image. Applied to the case of voice, the implementation of a system whose objective is to recognise an individual on the basis of his or her voice characteristics – this is called speaker recognition – is a form of biometric data processing. Over the years, the CNIL has developed a doctrine relating to the regulation of biometric data use, whether for fingerprint, face, iris or silhouette recognition, or even speaker recognition[13]. For example, several trials have been authorised by the CNIL for the voice authentication of retail bank users on interactive voice servers[14]. To this end, technical and organisational measures had to be implemented in order to meet data protection requirements and in particular to guarantee that the persons using them have control over their biometric data.

[11] - Sonja Trent-Brown, *Voice quality: Speaker identification across age, gender, and ethnicity,* The Journal of the Acoustical Society of America, march 2018

[12] - Laetitia Jeancolas et al., *L'analyse de la voix comme outil de diagnostic précoce de la maladie de Parkinson : état de l'art, Compressions et Représentation des Signaux Audiovisuel,* may 2016

[13] - CNIL, Biométrie, https://www.cnil.fr/fr/biometrie

[14] - CNIL, *La CNIL autorise l'expérimentation de dispositifs biométriques de reconnaissance vocale par des établissements bancaires*, https://www.cnil.fr/fr/la-cnil-autorise-lexperimentation-de-dispositifs-biometriques-de-reconnaissance-vocale-par-des

A question to... Chloé Clavel

How in practice do we encode emotions and ensure their correspondence with a physical phenomenon?

What is the emotional phenomenon? The answer to this question is a controversial issue. Equipping the machine with the ability to understand human behaviour: this is the scientific challenge around which different scientific communities are gathering (signal processing, automatic language processing, artificial intelligence, robotics, human-machine interaction, etc.). The types of information available are the signals acquired by the system via sensors (image, sound, physiological sensors). The data manipulated is therefore of very low level: sound samples or image pixels. In the example of the voice, many of the acoustic descriptors used to characterise the different emotional states are intended to model changes in the acoustic signal related to physiological changes at the base of the glottis. The bodily or physiological changes that accompany certain emotional states will strongly influence how the speaker's oral message is produced. For example, in the case of fear, the typical physiological changes are increased pulse rate and blood pressure and dry mouth, manifested by a louder, higher-pitched voice and a faster flow, as opposed to boredom and sadness, which are correlated with a lower heart rate and manifest as a lower, less intense voice and slower flow.

Chloé Clavel is an associate professor in Affective Computing at Telecom Paris

> Full interview to be found on LINC [in French]
Chloé Clavel, *Les machines ne font « pas encore » mieux que les humains pour interpréter les émotions*, Linc.cnil.fr, october 2018, https://linc.cnil.fr/itw-chloe-clavel-les-machines-ne-font-pas-encore-mieux-que-les-humains-pour-interpreter-les-emotions

Voice can also be used to infer very private information without it being considered sensitive by the regulations. For example, more and more companies are offering to analyse the voice signal in order to extract information about an individual's emotional state. A wide range of purposes are cited: to allow telephone advisers to know and analyse the mood of their interlocutors in real time, to ensure their professionalism and the due performance of the job, to analyse the postures, attitudes and inter-personal skills of candidates who have recorded cover letters and video CVs[15], all applications whose implementation raises many questions[16]. Thus, although such data does not have a special status in the data protection rules, its use is nevertheless likely to spark a sense of privacy invasion among the data subjects[17].

[15] - Anne Rodier, *Le robot, fidèle compagnon du recruteur,* Le Monde, february 2020, https://www.lemonde.fr/economie/article/2020/02/22/le-robot-fidele-compagnon-du-recruteur_6030459_3234.html

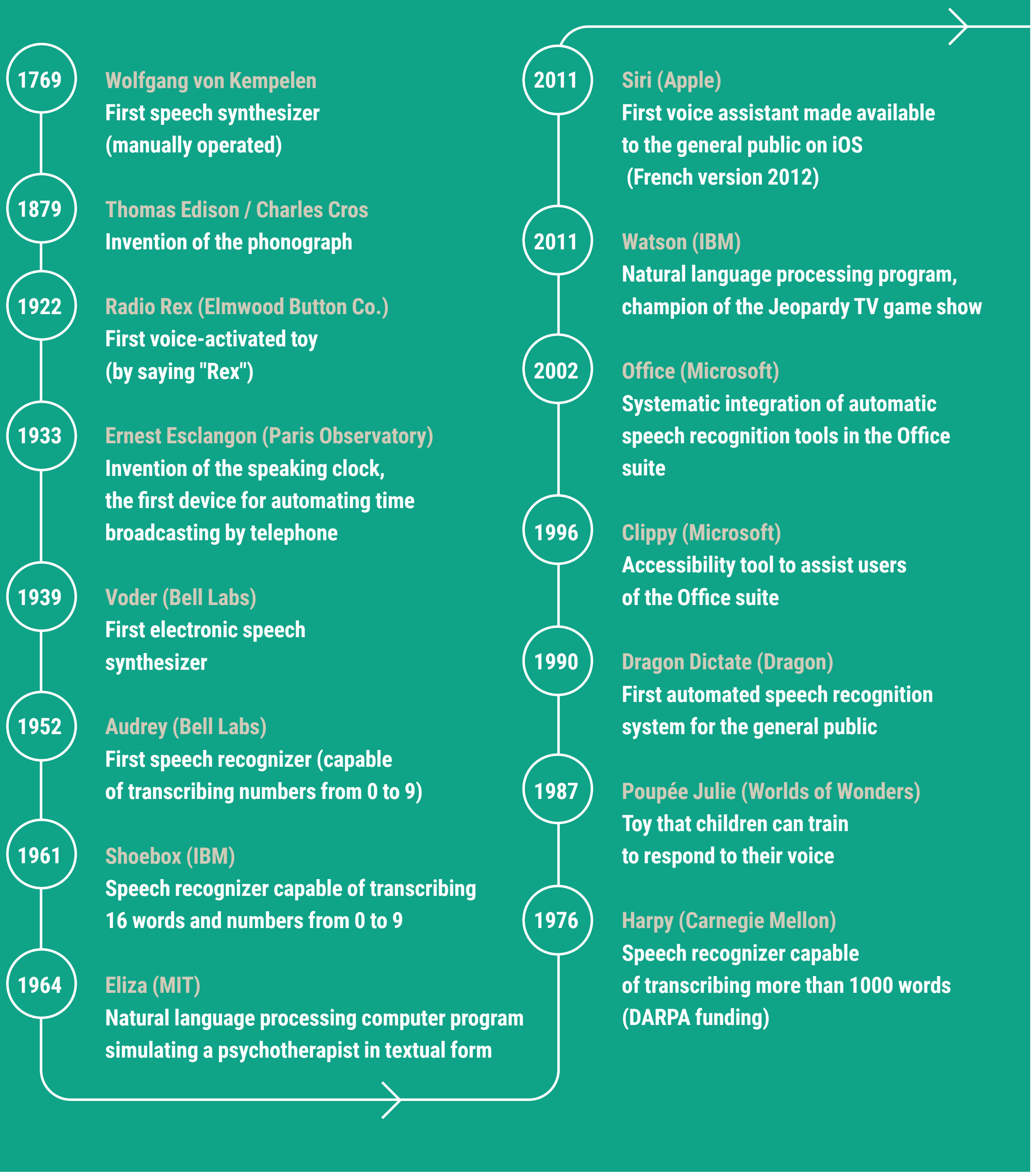
[16] - Laurence Devillers, *Les robots émotionnels: santé, surveillance, sexualité... et l'éthique dans tout cela ?*, L'observatoire, 2020.

[17] - Régis Chatellier, *Captation des émotions : comment vous le direz pourra être retenu contre vous...*, Linc.cnil.fr, april 2018, https://linc.cnil.fr/fr/captation-des-emotions-comment-vous-le-direz-pourra-etre-retenu-contre-vous

VOICE ASSISTANT, WHO ARE YOU?

While voice processing issues are not new, the gradual arrival of voice assistants within our personal equipments casts them in a fresh light. To fully understand the issues raised by these new devices, it is essential to understand where they come from and how they work.

History of Voice Assistants and their Ancestors



1769 Wolfgang von Kempelen
First speech synthesizer (manually operated)

1879 Thomas Edison / Charles Cros
Invention of the phonograph

1922 Radio Rex (Elmwood Button Co.)
First voice-activated toy (by saying "Rex")

1933 Ernest Esclangon (Paris Observatory)
Invention of the speaking clock, the first device for automating time broadcasting by telephone

1939 Voder (Bell Labs)
First electronic speech synthesizer

1952 Audrey (Bell Labs)
First speech recognizer (capable of transcribing numbers from 0 to 9)

1961 Shoebox (IBM)
Speech recognizer capable of transcribing 16 words and numbers from 0 to 9

1964 Eliza (MIT)
Natural language processing computer program simulating a psychotherapist in textual form

2011 Siri (Apple)
First voice assistant made available to the general public on iOS (French version 2012)

2011 Watson (IBM)
Natural language processing program, champion of the Jeopardy TV game show

2002 Office (Microsoft)
Systematic integration of automatic speech recognition tools in the Office suite

1996 Clippy (Microsoft)
Accessibility tool to assist users of the Office suite

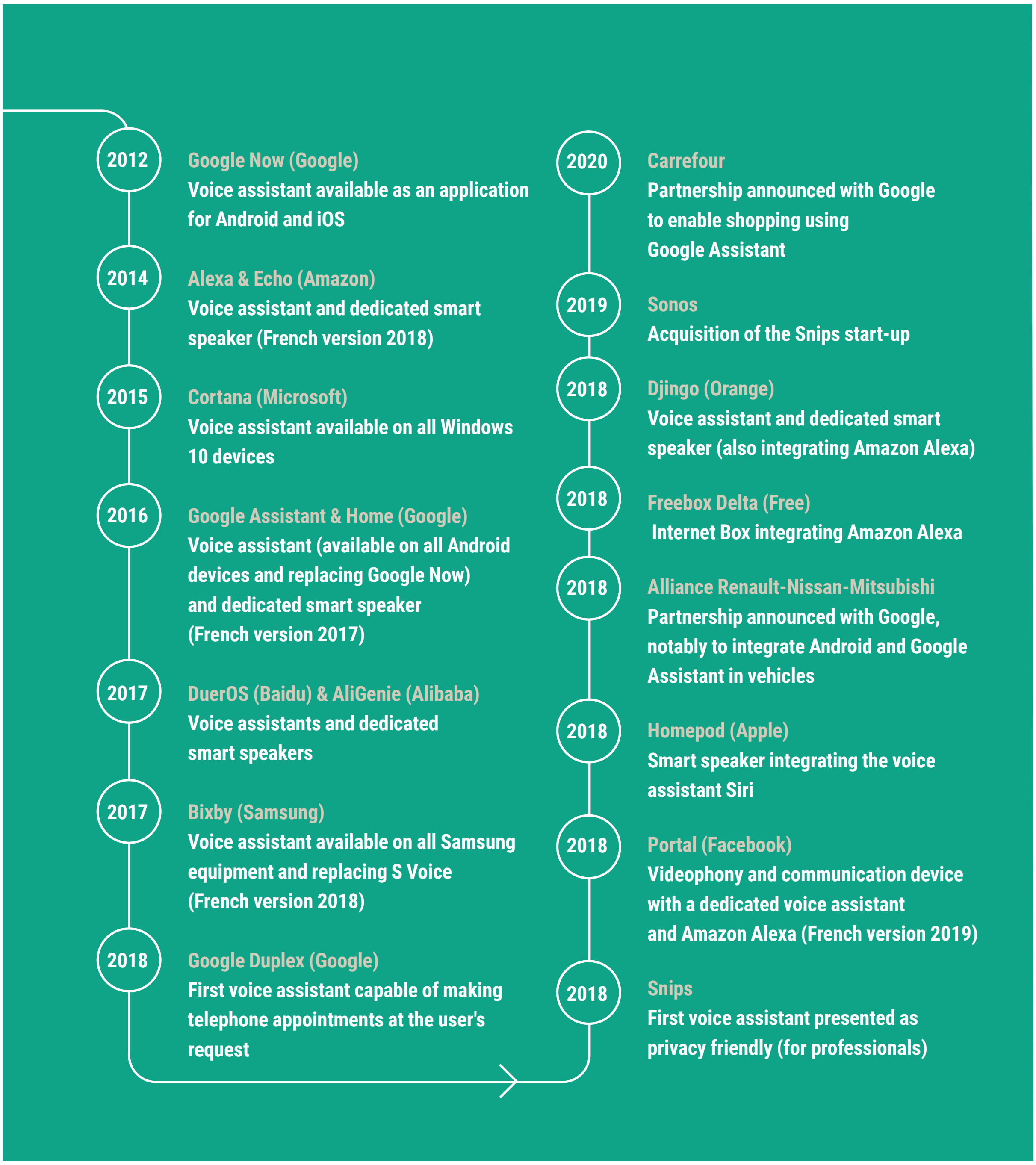
1990 Dragon Dictate (Dragon)
First automated speech recognition system for the general public

1987 Poupée Julie (Worlds of Wonders)
Toy that children can train to respond to their voice

1976 Harpy (Carnegie Mellon)
Speech recognizer capable of transcribing more than 1000 words (DARPA funding)

An approach that is already old...

It is only since the years 2010 that voice assistants have burst fully into the collective imagination. However, many steps preceded their entry on the scene. Indeed, no digital sound recording, automatic speech recognition, natural language understanding or speech synthesis capabilities – no voice assistant...



2012 **Google Now (Google)**
Voice assistant available as an application for Android and iOS

2014 **Alexa & Echo (Amazon)**
Voice assistant and dedicated smart speaker (French version 2018)

2015 **Cortana (Microsoft)**
Voice assistant available on all Windows 10 devices

2016 **Google Assistant & Home (Google)**
Voice assistant (available on all Android devices and replacing Google Now) and dedicated smart speaker (French version 2017)

2017 **DuerOS (Baidu) & AliGenie (Alibaba)**
Voice assistants and dedicated smart speakers

2017 **Bixby (Samsung)**
Voice assistant available on all Samsung equipment and replacing S Voice (French version 2018)

2018 **Google Duplex (Google)**
First voice assistant capable of making telephone appointments at the user's request

2020 **Carrefour**
Partnership announced with Google to enable shopping using Google Assistant

2019 **Sonos**
Acquisition of the Snips start-up

2018 **Djingo (Orange)**
Voice assistant and dedicated smart speaker (also integrating Amazon Alexa)

2018 **Freebox Delta (Free)**
Internet Box integrating Amazon Alexa

2018 **Alliance Renault-Nissan-Mitsubishi**
Partnership announced with Google, notably to integrate Android and Google Assistant in vehicles

2018 **Homepod (Apple)**
Smart speaker integrating the voice assistant Siri

2018 **Portal (Facebook)**
Videophony and communication device with a dedicated voice assistant and Amazon Alexa (French version 2019)

2018 **Snips**
First voice assistant presented as privacy friendly (for professionals)

[18] - Wolfgang Minker et Françoise Néel, *Développement des technologies vocales*, Le travail humain, 2002, https://www.cairn.info/revue-le-travail-humain-2002-3-page-261.htm

FOCUS ON...

Talking machines and fiction

Whether in literature or in film, the role of the speaking assistant has assumed different facets. Two main types seem to stand out. Fiction has made it possible to make it either a trusted partner, or on the contrary, an adversary with a cold intelligence, responsible for taking radical decisions against humans, who are by nature irrational. In the first case, we find a talking machine that plays the role of a side-kick to the point of being a protector: such as J.A.R.V.I.S., which allows Iron Man not to die in every movie (Jon Favreau, 2008 and following), GERTY in the film *Moon* (Duncan Jones, 2009) or the KITT car in the *K2000 series* (Glen A. Larson, 1982-1986). On the other hand, there is no shortage of examples of cold and ruthless calculating assistants. The most famous of them is of course HAL 9000, in *2001, the Space Odyssey* (Stanley Kubrick, 1968) but the *Tron* Master Control Program (Steven Lisberger, 1982) or V.I.K.I. of *I, Robot* (Alex Proyas, 2004) can also be cited. On a slightly different note, the question of the man-machine relationship is questioned in a more individual and social way, as for example in the film *Her* (Spike Jonze, 2013), in which assistant Samantha turns Joaquin Phoenix's head. The machine then becomes an ideal lover, a perfect partner or a close friend as illustrated in the series *Mr. Robot* (Sam Esmail, 2015), where FBI agent Dominique DiPierro shares his most personal thoughts with his voice assistant Alexa.

Fiction also leads to reflections on the notion of housing of the future, and on the roles of domestic assistants. The *Years and Years* series (Russell T. Davies, 2019) shows the place occupied by assistants in the years to come, making them one of the main channels for exchanges in the family. This device introduces new ways of communicating, calling each other, sending messages, etc. In an older version, a more dystopian side can be found in *Back to the Future 2* (Robert Zemeckis, 1986), a film in which the connected house is the only one to respond to Marty McFly (Michael J. Fox) when he comes home from work, greeting him with nicknames like "lord of the manor" or "king of the castle". We could even go all the way back to *Snow White* (Jacob and Wilhelm Grimm, 1812) and the witch's magic mirror, which answers her questions after the words "Mirror, mirror" have been uttered.

Other works may also be cited, without this list being exhaustive: *The Outer Zone* (Alain Damasio, 1999), WOPR in *WarGames* (John Badham, 1983), Data in *Star Trek* (Gene Roddenberry, 1966), *Blade Runner* (Ridley Scott, 1982), Alpha 60 *in Alphaville* (Jean-Luc Godard, 1965), Icarus in *Sunshine* (Danny Boyle, 2007), Auto in *Wall-E* (Andrew Stanton, 2008), etc.

What is a voice assistant?

A distinction is made between natural languages and formal languages. The semantics of a formal language are set to be unambiguous for the purpose of developing a computer program. A human operator who develops or wishes to interact with a computer program developed in formal language must therefore adapt to it, both in terms of lexicon and syntax. The semantics of natural language, on the other hand, are specific to human language. Depending on the characteristics of the language and the diversity of the lexicon, the same instruction may be formulated in multiple ways, while some commands may seem similar but refer to two different objects. Inference mechanisms are then frequently used to resolve these ambiguities, for example, as mentioned above, the time when the instruction has been stated, the place, the person's interests, etc.

> «
> ***A voice assistant can be as a software application that provides capabilities for oral dialogue with a user in natural language.***
> »

A voice assistant can be broken down into modules to perform different tasks: sound capture and restitution, automatic speech recognition (speech to text), natural language processing, dialogue strategies, access to ontologies (data sets and structured concepts related to a given domain) and external knowledge sources, language generation, text-to-speech synthesis, etc. In concrete terms, the assistant must allow interaction in order to carry out actions ("turn on the radio", "turn off the light") or to access knowledge ("what will the weather be like tomorrow?", "is the 7:43 train running?"). It thus plays the role of intermediary and orchestrator which is supposed to facilitate the user's tasks.

In practice:

A voice assistant is not a smart speaker... but a smart speaker can be equipped with a voice assistant

It is common to confuse a voice assistant with a smart speaker. However, the second is only a material embodiment or form factor of the first. A voice assistant can be deployed in a smartphone, a smart speaker, a connected watch, a vehicle, household appliances, etc.

A voice assistant is a personal assistant... but the reverse is not necessarily true

A personal assistant is a software agent designed to interact in natural language with an individual to help him/her perform tasks. However, voice is not necessarily the interaction modality used. It can thus be a text-based exchange, as for example in the case of a chatbot.

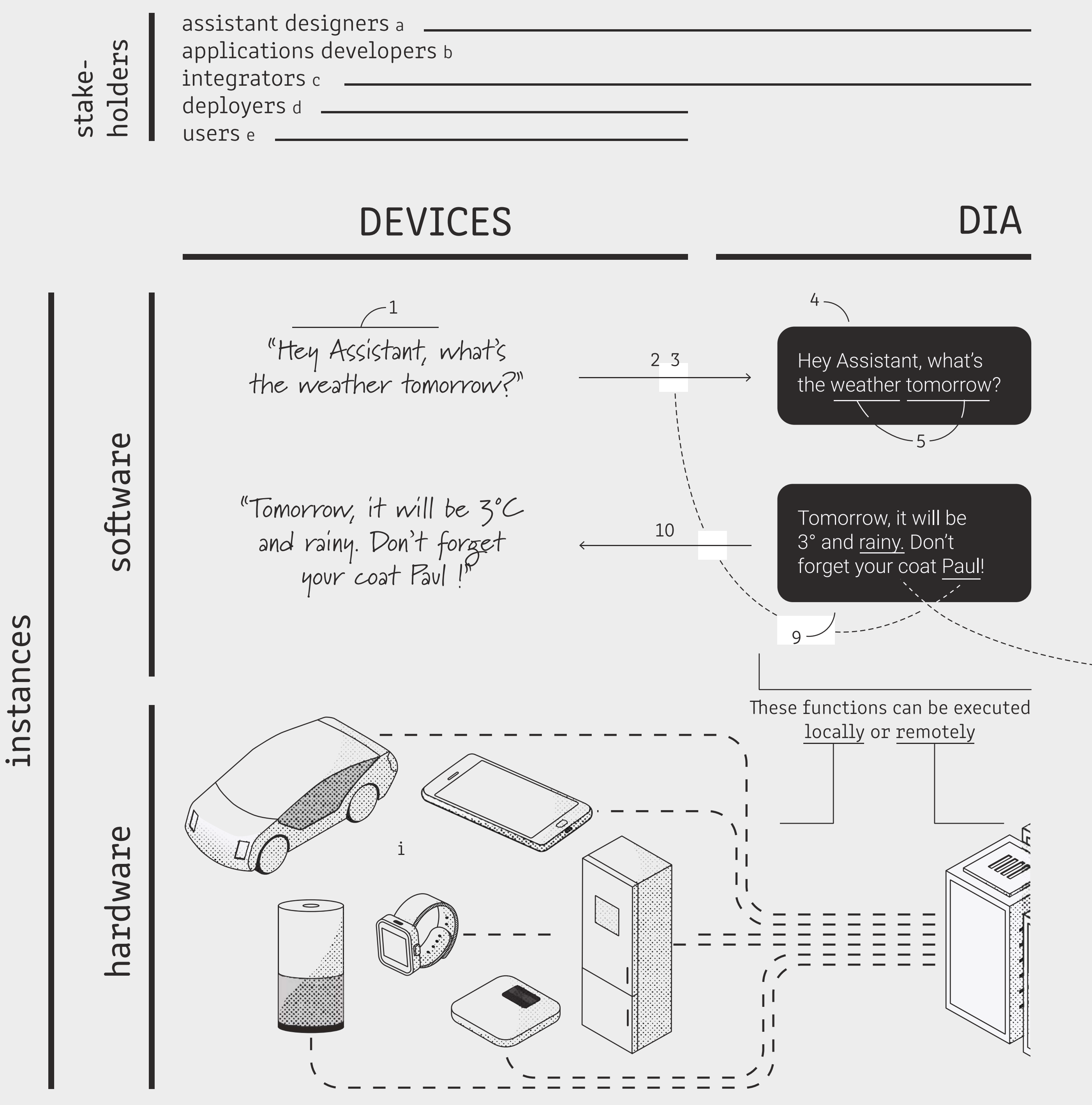
If the interaction with the voice assistant takes the form of a user's oral exchange with electronic equipment, in practice the organisation of the underlying data processing may involve multiple information flow patterns.

It is possible to isolate three main entities to understand how they work (see infographics on page 14):

- **the physical entity:** hardware element in which the assistant is embodied (smartphone, speaker, TV, etc.) and which includes microphones, loudspeakers and computing capabilities (more or less developed depending on the case);

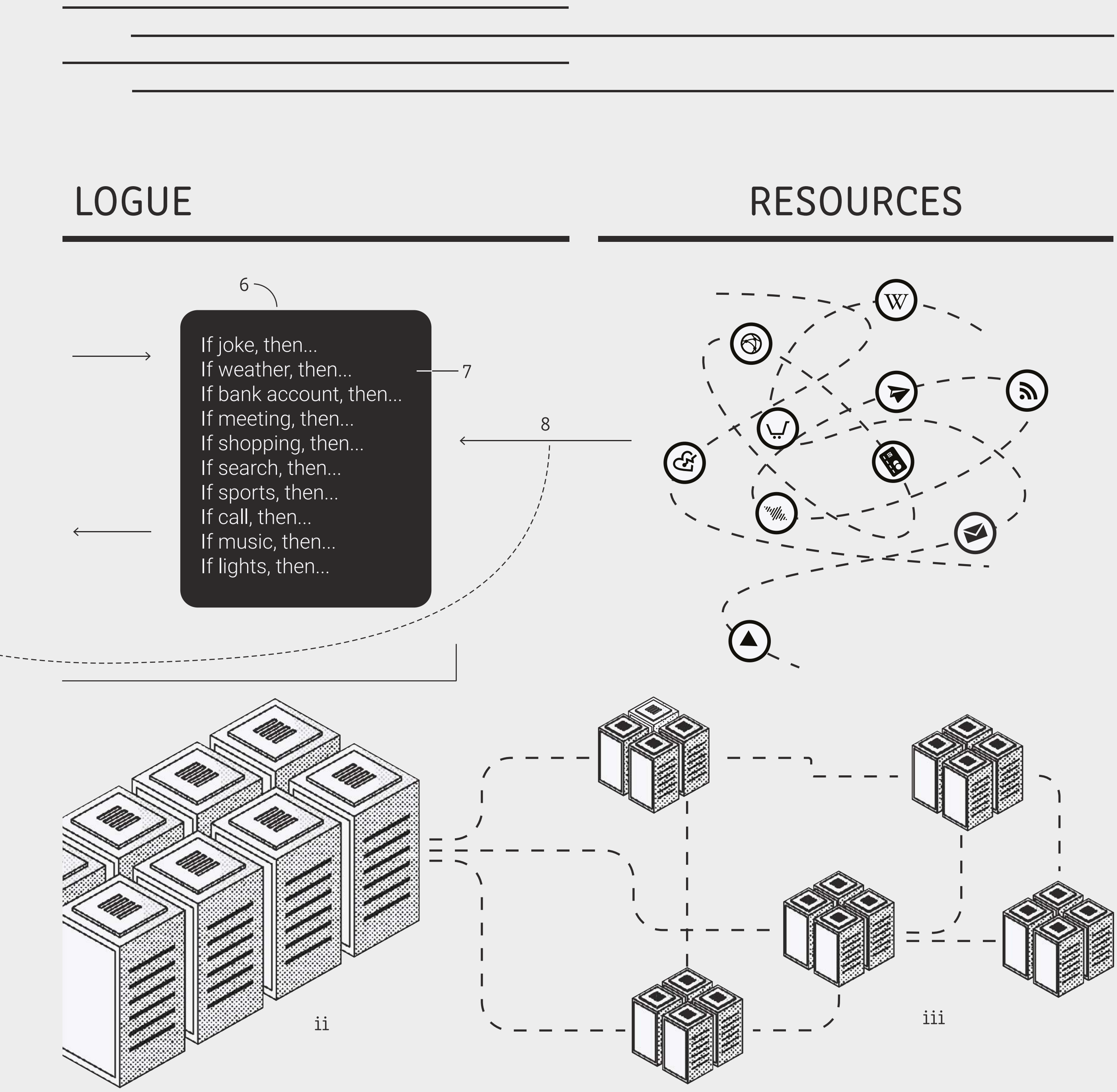
- **the software entity:** the component implementing human-machine interaction as such and which includes built-in modules for automatic speech recognition, natural language comprehension and generation, dialogue and speech synthesis. This can be done directly inside the hardware element, but is in many cases done remotely;

- **resources:** external data such as knowledge bases, ontologies or business applications that provide knowledge ("indicate the time on the West Coast of the United States") or enable the requested action to be carried out in practice ("increase the temperature by 1.5°C").


stake-holders
assistant designers a
applications developers b
integrators c
deployers d
users e
DEVICES
DIA
instances
software
hardware
1
"Hey Assistant, what's the weather tomorrow?"
2 3
4
Hey Assistant, what's the weather tomorrow?
5
"Tomorrow, it will be 3°C and rainy. Don't forget your coat Paul !"
10
Tomorrow, it will be 3° and rainy. Don't forget your coat Paul!
9
These functions can be executed locally or remotely
i

THE DATA PATHWAY

An infographics to understand the inner workings of voice assistants.



If joke, then...
If weather, then...
If bank account, then...
If meeting, then...
If shopping, then...
If search, then...
If sports, then...
If call, then...
If music, then...
If lights, then...

captions

Stakeholders

a. Develop the voice assistant operating modes.
b. Develop applications to be deployed on the assistant.
c. Integrate the assistant in their devices.
d. Install the assistant in spaces under their responsibility.
e. Use a device embedding a voice assistant.

Software

1. Local wake word detection
2. Wake word check
3. Speaker recognition
4. Automatic speech recognition
5. Intent detection
6. Dialogue management
7. Intent selection
8. Informations retrieved in public resources or accessible through authentication
9. Natural language generation
10. Speech synthesis

Hardware

i. Household devices
ii. Servers of the voice assistant designers
iii. Servers of the application developpers.

How does a voice assistant work?

As shown in the infographics on page 14, a series of tasks has to be performed to carry out an action or to access information.

0) Deployed within a piece of equipment (smartphone, smart speaker, vehicle), the voice assistant is on standby. In concrete terms, it is constantly listening, while possibly concentrating its listening on certain areas of space, for example to try to neutralise sound sources such as a television set (using the spatial filtering techniques such as beam forming). However, it does not store the audio data and does not perform any operations until a specific wake word has been heard. For this purpose, a buffer of a few seconds is used.

1) When the user utters the wake word, the assistant "wakes up". A listening channel opens and the audio content is streamed.

2) In many cases, if the processing is done remotely, a second check of the wake word pronunciation is performed on the server side in order to limit unwanted triggering.

3) It is possible for the user, if s/he has been previously enrolled – that is to say, if his/her vocal characteristics have been learned from voice samples s/he has produced – that the speaker is identified (speaker recognition).

4) The user shall state his/her request, which is forwarded to the processing entities. These may be remote servers, or in the case of local processing, hardware resources embedded in the equipment. The spoken sequence of speech is then automatically transcribed (speech-to-text).

5) Using natural language processing (NLP) technology, speech is interpreted. The intentions of the message are extracted and the information variables (slots) identified.

6) A dialog manager clarifies the interaction scenario to be implemented with the user by providing the appropriate response scheme.

7) An appropriate response to the user's request is identified and, if necessary, remote resources are used. They can be publicly accessible knowledge database (online encyclopaedia, etc.) or resources accessed by authentication (bank account, music application, customer account for online purchase, etc.).

8) The slots are filled with the knowledge retrieved.

9) An answer phrase is created and/or an action is identified (raise the blinds, increase the temperature, play a piece of music, answer a question, etc.).

10) This sentence is synthesised (text-to-speech) and/or the action to be performed is sent to the equipment.

11) The response and/or command is implemented by the voice assistant-enabled equipment.

12) The voice assistant returns to standby.

A voice assistant or a speaker is therefore not "smart" strictly speaking. The knowledge items come from third party sources: freely accessible data (online encyclopaedias), databases containing information filled in by the user (his/her diary, address book, etc.), etc.

«

A voice assistant or a speaker is therefore not "smart" strictly speaking.

»

FOCUS ON...

The development of speech processing technologies

Following the establishment of the theoretical bases of signal processing, notably Claude Shannon's theories of information and sampling, automatic speech processing has become a core module of engineering science. At the crossroads of physics (acoustics, wave propagation), applied mathematics (modelling, statistics), computer science (algorithms, learning techniques) and humanities (perception, reasoning), speech processing has rapidly become a subject of study: speaker identification and verification, automatic speech recognition, speech synthesis, emotion detection, etc. Over the last 15 years or so, the discipline as a whole has made very significant progress, with various factors contributing to this: improved methods, a significant increase in computing capacity and greater volumes of data available. The excellence of French research in this field is also evident with historically recognised laboratories and centres such as, among others, LIMSI (Paris Saclay), LIUM (Le Mans), LIA (Avignon), LORIA (Nancy), LIG et GIPSA-lab (Grenoble), LPL (Aix-en-Provence), IRIT (Toulouse), Eurecom (Sophia-Antipolis), Ircam, LPP (Paris), etc. (more information available on the website of AFCP, the French language association for spoken communication[19]).

Automatic speech recognition used to involve three distinct steps to: 1) determine which phonemes had been pronounced using an acoustic model; 2) determine which words were pronounced using a phonetic dictionary; 3) transcribe the sequence of words (sentence) most likely to have been spoken using a language model. Today, with the progress made possible by deep learning (a machine learning technique), a large number of systems offer automatic transcription of speech from end to end. This eliminates the need for complex training of three different models while providing better performance in terms of both results and processing time. Almost all the leading tech companies now offer their own variants that can be easily used by API systems, but open-source systems also exist (DeepSpeech[20] or Kaldi[21] for example).

Since the 1990s, speech synthesis has been used mainly in the form of a synthesis known as concatenation of units. This technique consists in selecting – from all the recordings of an actor previously transcribed into phonemes, syllables and words – the sound blocks that correspond to the words we want the voice to pronounce and assembling them one after the other to form an intelligible sentence with natural diction. The advantage of this synthesis is that it is exclusively based on the reuse of real blocks and therefore guarantees the naturalness of the synthetic voice. Its disadvantage, however, is that it is limited to the person's voice and its stylistic and expressive content. Statistical or parametric synthesis appeared in the late 1990s with the first attempts to model the parameters of a voice such as intonation, rhythm, and timbre, using generative statistical models such as hidden Markov chains. If synthesis by concatenation is still widely used, the leading tech players are now dominating research and development in this sector with achievements focusing on parametric synthesis such as WaveNet[22], Tacotron[23], DeepVoice[24], or the demonstration of Google Duplex[25] where the synthesised voice makes a hair appointment.

[19] - http://www.afcp-parole.org/

[20] - https://github.com/mozilla/DeepSpeech

[21] - https://github.com/kaldi-asr/kaldi

[22] - Aäron van den Oord et Sander Dieleman, WaveNet: A generative model for raw audio, Deepmind blog, september 2016, https://deepmind.com/blog/article/wavenet-generative-model-raw-audio

[23] - Yuxuan Wang, Expressive Speech Synthesis with Tacotron, Google AI blog, march 2018, https://ai.googleblog.com/2018/03/expressive-speech-synthesis-with.html

[24] - Deep Voice 3: 2000-Speaker Neural Text-to-Speech, Baidu Research blog, october 2017 http://research.baidu.com/Blog/index-view?id=91

[25] - Yaniv Leviathan, *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*, Google AI blog, may 2018, https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

FOCUS ON...

The development of natural language processing technologies

Natural Language Processing (NLP) is a multidisciplinary field involving linguistics, computer science and artificial intelligence. It aims to create natural language processing tools for various applications: machine translation, automatic text generation and summarisation, spelling correction, question answering systems, text mining, named entity recognition, sentiment analysis, etc. The concept took off in the 1950s with the work of Alan Turing and his famous test to characterise whether in an exchange of written messages, a human subject can determine whether or not s/he is addressing a machine[26].

Concretely, the objective of NLP is to enable machines to give meaning to human exchanges. This is a significant challenge because computer tools traditionally require interaction with them in a formal programming language, i.e. precise, unambiguous and highly structured. However, human speech can be imprecise, equivocal and its structure can vary depending on the level of language used, field of application, etc. The two main branches of NLP are syntactic analysis and semantic analysis. The first one allows the meaning of a sentence to be analysed according to grammatical rules. It is thus a matter of relying on the arrangement of the words that constitute it. Among the techniques used are word segmentation (which divides a text into units), sentence breaks (which position the limits of a text's sentence), morphological segmentation (which divides words according to their composition) and stemming (which groups together words with a common root). Semantic analysis uses the meaning of the constituent elements of a text. Word-sense disambiguation (which derives the meaning of a word according to context) and named entity recognition (which consists of searching for textual objects such as proper nouns, dates, places, etc.) are frequently used techniques.

While early approaches to NLP were grounded in the use of artificial intelligence algorithms based on rule-based systems and ontologies, current approaches employ machine learning methods and more specifically deep learning. Here again, the improvements in the methods developed, the increase in available data and the development of computational capacities have made it possible to implement approaches based on machine learning to make statistical inferences from the analysis of very large corpora of texts. French laboratories and research centres working on the themes of natural language processing are often also involved in speech processing. The same actors are therefore involved as those indicated in the previous box. However, several other research centres come to mind, without the list being exhaustive: LS2N (Nantes), LIS (Marseille), IRISA (Rennes), GREYC (Caen), LIRMM (Montpellier), etc. (more information available on the website of ATALA, the association for automatic language processing[27]).

Who is behind a voice assistant?

The voice assistant involves five categories of actors (see infographics page 14):

- **The designers:** responsible for the development of the voice assistant, they design and define its functioning and possibilities: activation modalities, choice of architecture, data access, recordings management, hardware specifications, etc.
- **The application developers:** in the same way as they do for mobile applications, they want to develop a third-party application for a voice assistant. For this, it is necessary to comply with the development guidelines laid down by the designer.
- **The integrators:** manufacturers of connected objects and equipment, they wish to equip these with a voice assistant. To this end, they make sure that the minimum specifications defined by the designers are duly met.

[26] - Alan Turing, *Computing machinery and intelligence*, Oxford University Press, vol. 59, no 236, 1950
[27] - https://www.atala.org/

FOCUS ON...

Data, a key driver for the development of a voice assistant

Multiple technological building blocks are necessary for the proper functioning of a voice assistant. Traditionally, French and international scientific research has relied on the production of numerous corpora in order to develop and measure advances in speaker recognition, automatic speech recognition, keyword detection, etc. The National Institute of Standards and Technology in the United States (NIST) and the French Government Defence procurement and technology agency (DGA) have initiated extensive evaluation campaigns with the Speech Analytics programs[28] and Speaker and Language Recognition[29] for the first and the ESTER, ETAPE or REPERE campaigns for the second. Besides, since the 1990s, several organisations such as the European Language Resources Association[30] or the Linguistic Data Consortium have proposed to collect and distribute oral, written and terminological language resources to support the development of automatic voice and speech processing technologies.

Significant progress has been made, finally paving the way, from 2010 onwards, to the development of voice assistants for the general public. However, with a few exceptions (such as the BERT language model[32] made available by Google and for which there are French versions such as CamemBERT[33]), most of the data used by the predominant tech giants is not readily available. However, several initiatives can be highlighted such as the Common Voice project[34] initiated by Mozilla that aims at collecting audio recordings in many languages, or the Voice Lab[35], a French association bringing together some thirty players whose aim is also to build up vocal resources that take into account accents as well as regional and international dialects. Linguistic diversity, particularly for inclusion purposes, must indeed be taken into account, whereas the major market players generally focus on those languages, dialects and accents considered to be the most profitable. Also, worth noting, the public authorities themselves have taken up the issue of building language resources with the PIAF project[36].

- **The deployers:** in charge of places receiving the public (accommodation, professional environments, rental vehicles, etc.) they wish to provide their audience with voice assistants (possibly with dedicated applications).

- **The users:** As the final link in the voice assistant value chain, they may use voice assistants on various equipment (smart speaker, TV, smartphone, smart watch, etc.).

Depending on the business models (see the actors' strategies on page 24) and technological choices, some actors may assume several combinations of roles, for example, designer and integrator, application designer and developer, etc.

A voice assistant, what for?

There are several advantages to using a voice interface, such as: the naturalness of the interaction which does not involve specific learning, the speed of execution of the command which is close to the performance of an expert in keyboard typing and the extension of the field of action which can allow faster access to information. While relying on speech also brings with it difficulties in interpreting the message correctly (variability of the audio signal between different speakers, the acoustic environment, ambiguity of the language, etc.), its benefits have also been clearly identified.

[28] - https://www.nist.gov/programs-projects/speech-analytics
[29] - https://www.nist.gov/programs-projects/speaker-and-language-recognition
[30] - http://www.elra.info/en/
[31] - https://www.ldc.upenn.edu/
[32] - Jacob Devlin et Ming-Wei Chang, *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*, Google AI blog, november 2018, https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html
[33] - https://camembert-model.fr/
[34] - https://voice.mozilla.org/fr
[35] - http://www.levoicelab.org/
[36] - https://piaf.etalab.studio/

In practice, more fluid or simpler task-making remains the primary reason for equipping oneself with voice assistants. This may be, for example, to make/answer a call, start a timer, etc., especially when the user's hands are unavailable because s/he is doing the washing-up,getting dressed or doing odd jobs. The connected home and home automation are the main uses promoted by the designers of voice assistants. By proposing to simplify the execution of tasks (turning on the light, adjusting the heating, lowering the shutters, etc.) and to centralise them through a single tool that can be easily activated remotely without requiring an intermediary, they chime with the arguments as a domestic facilitator. In addition to personal or household use, the use of voice commands may be of interest in professional environments where it is difficult to handle computer tools and use written commands (e.g. manufacturing work).

In theory, the major beneficiaries of the voice interface could be disabled or dependent people for whom the use of traditional interfaces is problematic. In other words, voice assistance can enable easier access to information and computer resources and thus promote inclusive mindsets. Sociologist Dominique Pasquier points out in particular that going through the voice makes it possible to overcome the difficulties associated with the written word, which can be found among the working classes[37].

Finally, health is also an area where there are many potential uses for conversational agents, whether they are voice-enabled or not, as indicated by the Lab e-santé, a think tank specialising in digital health issues, in its analysis[38]. Although there are still few examples of use in France, for some, it is the entire patient care pathway that could, in the long term, be impacted by human/assistant interactions: not only for well-being and prevention, but also for curative and supportive treatment. Numerous partnerships have been forged between leading tech players and medical professionals (see box on page 26).

FOCUS ON...

Seniors, the primary beneficiaries of these new tools?

The latest demographic studies indicate that, if in 2020, people aged 65 years and over account for approximately 19.6% of the French population, this proportion is expected to increase sharply. According to INSEE estimates, it is projected to grow to about a quarter of the population in 2040[39]. This societal change introduces a number of challenges related to the specific needs of older people, whose advancing age is often accompanied by poorer health, difficulties in pursuing a social life and even, in some cases, in carrying out routine activities.

Voice assistants, whose use does not require the manipulation of a keyboard, screen or mouse, but, at least in principle, only the use of voice commands close to natural language, hold a promise of simplicity and come across as an aid to overcome certain disabilities, which has made them a popular focus for research in the field of the "silver economy". Several solutions are currently marketed in France. Examples include the OLGA smartphone (Flagtory SAS) and the SkipIt (HomeKeeper consortium) and Fanny (Dynséo) speakers. These devices, which vary in technical architecture, have an interface that includes a voice assistant. This allows basic functionality to be managed without physical contact with the device: write and listen to a message, make a call, turn on the radio and play music. Finally, many specialised modules that can be built into "classic" voice assistants (Siri, Alexa, Google Assistant, etc.) exist and target the needs of the elderly through applications such as memory games, reading with a synthesised voice, or calling emergency numbers.

[37] - Dominique Pasquier, *Dans les classes populaires, la vie privée relève moins de l'individu que du groupe familial,* Linc.cnil.fr, march 2020, https://linc.cnil.fr/dominique-pasquier-dans-les-classes-populaires-la-vie-privee-releve-moins-de-lindividu-que-du-groupe

[38] - E-Health Lab, *Chatbot : le futur de la santé passe-t-il par le conversationnel ?,* july 2019, https://www.ticsante.com/documents/201907041716270.Livre_blanc_chatbot_du_Lab-esante.pdf

[39] - INSEE, Tables of the French economy (2018 edition), sheet 3.2 "Population by age", https://www.insee.fr/fr/statistiques/3303333?sommaire=3353488

WHAT ARE THE VARIOUS USES OF VOICE ASSISTANTS?

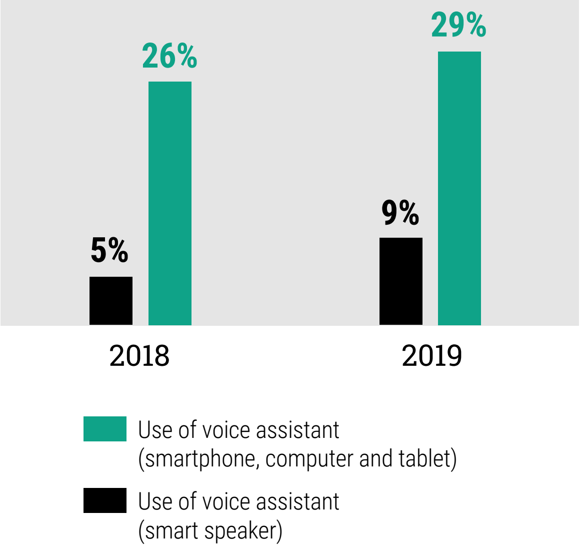
For designers of voice assistants, the promise of these devices is often to act as the butler of the connected home. However, current uses still seem to be in their infancy, and it is legitimate to wonder, quoting design professor Anthony Masure: "If voice assistants are the solution, what's the problem?"[40].

More and more devices...

For several months now, many figures and projections have been forecasting a growing distribution of these new digital tools. There is even talk of a trend comparable to the one witnessed for tablets. In the summer of 2019, there were reports of approximately 1.7 million voice assistants on connected speakers and 16 to 20 million users of the smartphone voice assistant in France. Worldwide, the consulting firm Roland Berger estimates the number of voice assistants, all media combined, at 3 billion, with a projection of 8 billion by 2023[41]. These figures could progress further thanks to the ever-increasing integration of these assistants into everyday objects, starting with the connected car.

Since 2018, the CNIL has included a question on voice assistants in the annual survey carried out with Médiamétrie on the digital practices of the French population[42] [43]. In light of the answers provided by the 2,000 or so people interviewed, several lessons can be drawn. First of all, usage is still fairly low with a third of respondents stating that they have used a voice assistant in the last 12 months. These indicators are nevertheless on the rise for 2019. Smartphone assistant usage increases from 26% to 29% when the use of a connected speakerphone voice assistant nearly doubles from 5% to 9% (see Figure 1). At the same time, 20% of Internet users surveyed said they had disabled the personal digital assistant (PDA) on their smartphone. Finally, uses are becoming clearer and the settings on these objects are increasingly being configured. 46% of respondents who have used this type of tool (i.e. around 700 people) indicate that they have already set up their voice assistant: checked the configuration, deleted the history of past voice commands, etc. This represents an increase of 6 points over what was measured the previous year.

Figure 1
Growth in the use of voice assistants in France between 2018 and 2019
(Base: Internet users aged 15 and over equipped with smartphone, computer or smart speaker (n=2,102))



[40] - Hubert Guillaud, « *Si les assistants vocaux sont la solution, quel est le problème ?* », InternetActu blog LeMonde.fr, january 2019, https://www.lemonde.fr/blog/internetactu/2019/01/13/si-les-assistants-vocaux-sont-la-solution-quel-est-le-probleme/

[41] - Cabinet Roland Berger, *La révolution naissante des assistants vocaux*, july 2019, https://www.rolandberger.com/fr/Publications/La-r%C3%A9volution-naissante-des-assistants-vocaux.html

[42] - LINC, *[Baromètre LINC 2019] - Les pratiques de protection des données progressent*, Linc.cnil.fr, december 2019 https://linc.cnil.fr/fr/barometre-linc-2019-les-pratiques-de-protection-des-donnees-progressent

[43] - LINC, *[Baromètre LINC 2018] - Des utilisateurs plus passifs vis-à-vis des assistants vocaux que des smartphones ou navigateurs*, Linc.cnil.fr, october 2018 https://linc.cnil.fr/fr/barometre-linc-des-utilisateurs-plus-passifs-vis-vis-des-assistants-vocaux-que-des-smartphones-ou

The CSA and Hadopi's May 2019 report *Assistants vocaux et enceintes connectées* also discusses the equipment requirements for people without smart speakers[44]. According to their estimates, only 4% of them would be willing to buy one. The reasons are first of all the "pointlessness" of these devices (67%), but also the security of personal data (59%). In one of its reports, Microsoft tends to confirm this issue[45]. It also highlights the questions relating to the security of the devices, the protection of personal data and the possibilities of espionage via passive listening. On the contrary, people who already have a voice assistant see it as a "real innovation" that will "revolutionise everyday life" (for 53% of them). Finally, ARCEP's annual barometer shows that the youngest generation is the most likely to appreciate this type of equipment, and that these assistants are more common in large households[46].

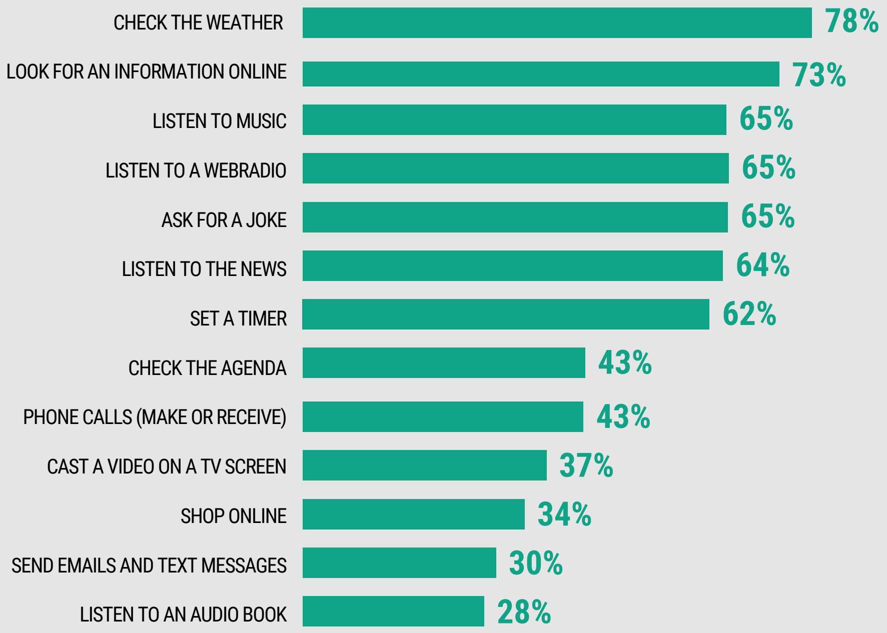
... for still cautious uses

According to the LINC 2019 Barometer, only 33% of Internet users have used a voice assistant in the last 12 months. Uses mainly revolve around "facilitating basic routine tasks". On this last point, the report published by PwC states that "the majority of items purchased are small and quick and are things that someone could buy without necessarily having to see it physically (to determine quality for example)"[47]. In concrete terms, when the assistant is used to make an expenditure, it is for small purchases, meal orders or groceries. The survey presented in the CSA-Hadopi report reinforces this idea of basic use (see Figure 2), and of more intrusive use being more seldom (purchases, dictation of text messages or emails, etc.). This impression is also confirmed by looking at which applications are most downloaded by users. In the case of Amazon Alexa, it is primarily radio content, games or sound environments (animal sounds or ocean sounds).

Figure 2

How are voice assistant-enabled speakers used

Source CSA-Hadopi report, users of smart speakers over the 30 days leading up to the survey, 287 individuals.



[44] - CSA-Hadopi, *Assistants vocaux et enceintes connectées - L'impact de la voix sur l'offre et les usages culturels et médias*, may 2019, https://www.csa.fr/Informer/Collections-du-CSA/Thema-Toutes-les-etudes-realisees-ou-co-realisees-par-le-CSA-sur-des-themes-specifiques/Les-autres-etudes/Etude-HADOPI-CSA-Assistants-vocaux-et-enceintes-connectees

[45] - *Microsoft, Voice report : From answers to action: customer adoption of voice technology and digital assistants*, 2019 https://advertiseonbing-blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voice-report/bingads_2019voicereport.pdf

[46] - ARCEP, *Le baromètre du numérique*, november 2019, https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/numerique/le-barometre-du-numerique.html

[47] - PwC, Report : *Prepare for the voice revolution*, february 2019, https://www.pwc.com/cisvoiceassistants

And tomorrow?

[Design Fiction]

CNIL	**2021**	Availability of an application Need help for the general public on major app stores
RATP	**2022**	Deployment of voice assistants at RATP bus stops to consult the traffic situation in real time and find out about your route
CPAM	**2023**	Launch of the operation "My GP is a chatbot" which offers a preliminary consultation with a voice assistant before any medical appointment is made
Pôle Emploi	**2024**	Implementation of voice assistants in Pôle Emploi (Employment) centres to facilitate the first registration of new applicants
Medical Robotics Japan*	**2027**	First clinical trial proving the positive effect of a dedicated voice assistant on the survival of elderly people cared for at home
Startup Dial'Og[48]*	**2027**	Raised $100 million for its opinion poll, "the ultimate weapon of elected officials who listen in on voters". A speech recognition system to revolutionise citizen participation.
MIT-Inria	**2029**	Design of an unambiguous wake word that can be used with all assistants to avoid false activations
European Commission	**2030**	Adoption of a regulation strictly governing the commands that can be performed by a voice assistant in a vehicle following a series of fatal accidents caused by a misunderstanding with the user
Mind Technologies*	**2046**	First industrial model for a "thought assistant" that can be controlled with the mind.

* This is a made-up name

[48] - LINC, IP7 Report, *Civic tech, Data & Demos*, « Quand notre voix est entendue », p. 30, december 2019. https://linc.cnil.fr/fr/civic-tech-donnees-et-demos-le-cahier-ip7-explore-les-liens-entre-democratie-et-technologies

WHAT STRATEGIES FOR VOICE ASSISTANT DESIGNERS?

As early as 2018, the LINC, the CNIL's Digital Innovation Laboratory, foresaw that the positioning of players in the voice assistant market would first of all involve the development of their pre-existing business models (advertising, online stores, app stores, etc.), with a market structured around the American, Korean and Chinese digital tech giants[49]. These major players are replicating or extending the platform strategy that has made them successful: they multiply free services (for individuals and companies alike) in order to capture as many users as possible through network effects. On the sidelines are more specialised players who try to distinguish themselves by their performance or more precise aims in terms of integrating their products.

Various strategies depending on the players

Depending on their traditional professions, voice assistants designers have different economic positions.

- **A leverage functionality for smartphone or software sales**

For smartphone manufacturers or operating system developers – be it Apple, Samsung, Huawei or Google – having an integrated voice assistant first and foremost provides a competitive advantage in the smartphone sales market. The launch of Siri has contributed to the success of the iPhone and iPad. Since then, voice assistants have multiplied: Google Assistant, Samsung Bixby, Microsoft Cortana, etc. The interest first of all lies in deploying, in a kind of extension of the hands-free kit, the use of the smartphone. This feature has already gone beyond technological novelty and is now the basic offer and no longer a surprise to users. The presence of an assistant thus becomes a competitive advantage. Samsung released the English and French versions of Bixby in 2018 and Huawei announced that its assistant Xiaoyi would be deployed internationally. The Chinese manufacturer has even unveiled a name for the French version: Célia. There is therefore a strong trend towards market penetration and international product sales.

- **Providers of customers in the market places**

Leading platforms Google and Amazon are transforming their voice assistants into advanced bases for their price comparators and marketplaces. It is a new intermediary for their online sales activities, whether Google Shopping in the former case or the Amazon website in the latter. Thus, when a shopping order is placed via the assistant, it will initially be passed on in their online sales ecosystem. In particular, the assistant will make proposals from the choice referenced on its own site. The challenge for these players is to embed the act of purchasing as fully and seamlessly as possible into users' routines. This issue of seamless interaction is then shared with third party commercial application developers. This is referred to as v-commerce, for this new avatar of e-commerce available only through the voice interface. The position of intermediary putting an individual in contact with content providers has been the core business of large platforms such as Google and Amazon from the outset. The CSA-Hadopi report looks back in particular on this role occupied in the cultural industries sector, and the new relationship between content publishers and voice assistant designers: "Right from their launch, [Google, Amazon et Apple] sought to offer an enticing spread of cultural content and media by forging partnerships with publishers or promoting their own services".

[49] - Olivier Desbiey, *Ok Google et Siri ne suivent pas la même voie qu'Alexa ou Cortana*, Linc.cnil.fr, march 2018, https://linc.cnil.fr/fr/ok-google-et-siri-ne-suivent-pas-la-meme-voie-qualexa-ou-cortana

- **Data sources for profiling and targeted advertising**

To use smart assistants, in most cases it is necessary to spend some time configuring the settings, not least linking an account to the assistant. Designers can thus enhance the profile of their users through the use of the assistant, the applications (or skills) that are installed, the orders placed, etc. This is the classic application of the two-sided market model: free of charge on the user side, but paid by advertisers in order to refine their marketing capabilities[50]. They will then be able to target people who fit a certain type of profile that they feel is most likely to purchase their products. For example, in the case of Google, assistant interactions are a new source of information attached to a user account that also includes information about searches on the search engine, actions on Android phones, videos viewed on YouTube, or navigational routes on Maps.

- **Divergent strategies: Microsoft and Apple**

Even among the leading tech players, different positioning can be observed. Thus, the Cortana assistant is an example of Microsoft's approach: cater to professionals rather than individuals. The company has also announced that there will be no more Cortana applications on Android and iOS as of 31 January 2020[51]. However, the assistant is far from being abandoned. Cortana will be integrated into Microsoft's Office 365 services, as well as into the group's products, such as Outlook. Microsoft's strategy is to equip its products, thus making Cortana a full-fledged tool for the various Windows formats.

For its part, while Apple was a pioneer in the development and installation of its Siri voice assistant on all of its products, its smart speaker – the HomePod – came late to the market compared to the other major suppliers mentioned above (2018 in France and the United States). Yet another business mindset is at work here, with a product focused on music and sound quality and part of a closed ecosystem, just like other Apple products. The HomePod currently only works with Apple applications. Tim Cook's firm is therefore following a different strategy than that of mass distribution and easy access by its competitors.

- **White label solutions, another path for smaller actors**

Finally, some players choose to position themselves on B2B (business to business) models. The idea is to provide, based on "white label" rationales, a "custom-made" voice assistant to be built into the customer's equipment. This assistant is then marketed under the brand name of the buyer. This business model is often adopted by small(er) players. These have particularly included the French startup Snips, acquired by Sonos at the end of 2019, and Nuance Communications (a company that worked in partnership with Apple on the first versions of Siri) via its assistant Nina. This is also the case for Harman, a subsidiary of Samsung, which equips cruise ships with sound and light systems and has developed Zoe, a dedicated cruise assistant, installed in all the cabins of some MSC Cruises ships[52].

The new frontier of online players' platform strategy

The leading voice assistant manufacturers such as Google or Amazon have opted for a large-scale distribution strategy. To that end, they showcase the tens or hundreds of millions (or even billions!) of devices addressable by their assistant. In addition to this existing market clout, development has taken place in two ways: 1) the very easy roll-out of the assistant on any equipment equipped with a microphone, a speaker, a network interface and some computing resources, and 2) the ease of developing a third-party application, modelled on a mobile app store. The provision of documentation and a software development kit (SDK) allows any developer to easily join the assistant's ecosystem. However, while applications can be developed by anyone, they can only be built with the limited and standardised tools provided by the manufacturer. So far, this operating model is free of charge. The question of possible developments, for example based on more conventional market models, arises, however: is a payment system for ranking applications to be planned in the future?

[50] - Jean-Charles Rochet and Jean Tirole, *Plateform Competition in two-sided markets*, Journal of the European Economic Association, june 2003, https://www.tse-fr.eu/articles/platform-competition-two-sided-markets

[51] - Microsoft, *Changes to Cortana services*, november 2019, https://support.microsoft.com/en-gb/help/4531683

[52] - Harman, Real-time AI: Meet ZOE, *the cruise industry's world's first voice-activated digital assistant*, january 2019

FOCUS ON...

Diversification of uses through partnerships

As a result of sector-specific dissemination strategies, voice assistants are found in areas that have to do with our daily routines and privacy alike, such as health care. In this area, partnerships between designers of assistants and players in the sector – public as well as private – have emerged. One example is the partnership signed between Amazon and the UK's NHS (National Health Service). It raised criticism from civil society actors, calling for more transparency in the functioning of this partnership. For example, Privacy International has questioned the combination of health data from a public service and Amazon's business model based on data recovery for advertising purposes[53]. Amazon's assistant also supported the Alexa Diabetes Challenge partnership between Jeff Bezos' company and the pharmaceutical company Merck[54]. In this case, the aim was to encourage young startups to propose solutions to help or support diabetics through the Alexa voice assistant.

In the case of cultural players' content, the CSA-Hadopi report underlined Google's decision to scale up partnerships with the media. Agreements concluded "in particular with audio content publishers, in order to be able to offer information (partnerships with Europe 1, Radio France, RTL, L'Équipe, BFM Business [...])", aimed at promoting the products and services of the Mountain View company, which specifies, however, that it does not seek exclusivity or exclusion of other services. In a professional context, Microsoft forges partnerships, such as with the Chinese Xiaomi[55]. Similarly, Apple and Salesforce have joined forces to implement assistant tools such as Siri Shortcuts in the products developed by the CRM (customer relationship management) software publisher[56].

Partnership with large companies or institutions is therefore another means of dissemination, allowing assistants to be integrated into already existing and/or popular products. In France, a striking illustration of this is that of the operator and internet service provider (ISP) Free which, at the end of 2018, integrated the Alexa voice assistant directly into its Freebox Delta Internet box. The box now contains a microphone, a loudspeaker, some computational capabilities for wake word detection and also a physical button to mute the microphone. However, the increasing number of partnerships could raise issues of device neutrality. In a paper published in June 2018, the French regulatory authority for electronic communications, postal services and press distribution (ARCEP) analyses the role of these new intermediaries: "Smartphones, voice assistants, connected cars and other devices are proving to be the weak link in achieving an open Internet. Because they are not therefore neutral and may limit the freedom of users to choose content and services on the Internet"[57]. As such, this new entry point, through its closed and proprietary ecosystem operation, forces the user to access services according to rules set by the assistant provider.

[53] - Privacy International, *Alexa, what is hidden behind your contract with the NHS?*, december 2019, https://privacyinternational.org/node/3298

[54] - Alexa Diabetes Challenge, http://www.alexadiabeteschallenge.com/

[55] - LeBrief, Xiaomi and Microsoft join forces, *Cortana pourrait être utilisé dans une enceinte connectée du chinois*, Next Inpact, february 2018, https://www.nextinpact.com/brief/xiaomi-et-microsoft-s-associent--cortana-pourrait-etre-utilise-dans-une-enceinte-connectee-du-chinois-2824.htm

[56] - Apple Newsroom, *Apple and Salesforce combine the best business devices with the world's #1 CRM solution*, september 2018, https://www.apple.com/fr/ newsroom/2018/09/apple-and-salesforce-partner-to-help-redefine-customer-experiences-on-ios/

[57] - ARCEP, *L'influence des terminaux sur l'ouverture d'internet*, june 2019, https://www.arcep.fr/la-regulation/grands-dossiers-internet-et-numerique/linfluence-des-terminaux-sur-louverture-dinternet.html

> ***Smart home uses are still***
> ***relatively uncommon,***
> ***but considered to harbour potential***
> ***and undeniable***
> ***for users' day-to-day.***

(SOURCE CSA-HADOPI REPORT, ASSISTANTS VOCAUX ET ENCEINTES CONNECTÉES, MAY 2019)

In addition to smartphones and dedicated speakers, voice assistants are being increasingly embedded into household objects, in order to cater to more concrete and utilitarian uses and to enable their designers to enter new markets. Many examples already exist: Connected refrigerator Samsung Family Hub featuring the Bixby Voice Assistant[58], Nest Guard Thermostat (subsidiary of Alphabet/Google)[59], Lidl's food processor, Monsieur Cuisine connect[60], etc. The last two examples symbolise assistant designers' readiness to pollinate. They both caused a scandal because it turned out that they both included a microphone, even though this was not mentioned in the documentation that came with the products. Indeed, the latter was not "active", as the services offered by the two devices did not include voice commands... at the time of their sale. However, with a view to upgrading and becoming part of a broader service, manufacturers were ready to deploy a voice assistant when the time was right.

FOCUS ON...

The house that spied on me

In December 2017, the journalist Kashmir Hill chose to completely equip her house with connected objects. She wrote an article on this experience entitled *The House that Spied on Me*[61]. The LINC team then summarised the purpose of the experiment and its conclusions. The aim was to document and understand what happens when you decide to completely “smarten up” your home, and in particular with regard to data protection.

"In this smart home, each of the connected objects acted like a tracer installed on our browser, able, if we know how to use it, to deduce household habits and consumption patterns. This information could be of great interest to data brokers, marketing agencies and even other types of players, which is why it is necessary to support users in understanding how these objects work, and to ensure the implementation of means of information and collection of "free, informed and revocable" consent adapted to these interfaces".

> Article to be found on LINC [in French][62]

[58] - Samsung, *Family Hub*, https://www.samsung.com/fr/familyhub/

[59] - C. Scott Brown, *Nest Secure has an unlisted, disabled microphone*, Android Authority, february 2019, https://www.androidauthority.com/nest-secure-google-assistant-mic-950134/

[60] - Marie Turcan, Monsieur Cuisine Connect : *micro caché, Android non sécurisé… les dessous du robot cuiseur de Lidl*, Numerama, june 2019, https://www.numerama.com/tech/525214-monsieur-cuisine-connect-micro-cacheandroid-non-securise-les-dessous-du-robot-cuisine-de-lidl.html

[61] - Surya Mattu and Kashmir Hill, *The House That Spied on Me*, Gizmodo, february 2018, https://gizmodo.com/the-house-that-spied-on-me-1822429852?rev=1518027891546

[62] - Régis Chatellier, *L'espion qui me logeait* : assistants vocaux et objets connectés dans la maison, Linc.cnil.fr, april 2018, https://linc.cnil.fr/fr/lespion-qui-me-logeait-assistants-vocaux-et-objets-connectes-dans-la-maison

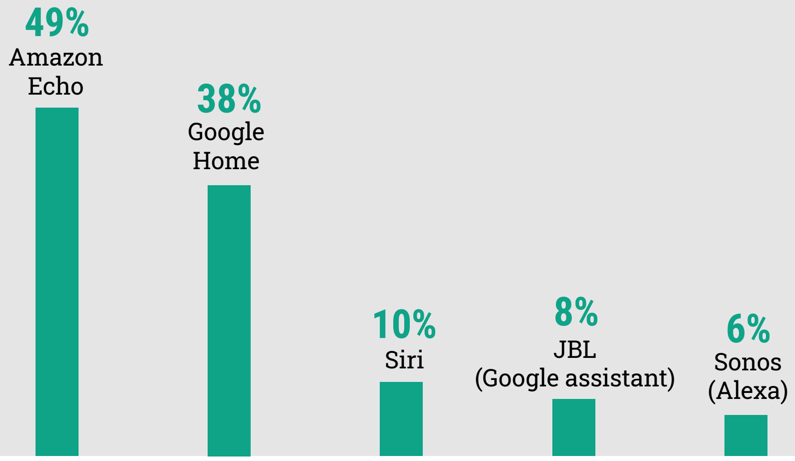
How is the market shared out?

In practice, the voice assistant market is strategic for online business and advertising models. The latter occupy a hegemonic position in the sale of voice assistants to the general public, while a host of smaller players are seeking to position themselves in highly targeted and specialised markets. Competitors have been waging a war of numbers since the beginning of 2019. If Amazon, for its part, claims to have passed the 100 million mark for the number of devices connected to Alexa[63], Google retorts with 1 billion objects which are Google Assistant-enabled[64]. However, these figures, which are still evolving, show a strong trend: while Amazon is behind Google and Apple in terms of devices with voice assistants, sales figures for 2019 indicate that the e-commerce company is the world's leading distributor of smart speakers. However, this observation must be put into context, as the CSA-Hadopi report reminds us: "While Amazon is heads and shoulders above the other market players in the US, its market share is more balanced with Google's at global level – a sign of the latter's strong position on export markets. Amazon dominates the smart speaker markets where its e-commerce business is well established [...]". If Facebook has never made a secret of wanting to join the race, Mark Zuckerberg's firm has not yet managed to launch its own product: his Aloha project has not (yet) seen the light of day[65], and its communication tool Portal relies for the moment on Amazon Alexa[66]. Finally, Asian manufacturers such as Baidu, Xiaomi, Alibaba and Tencent are making it increasingly clear they wish to position themselves on the Western market and compete with established players. Specifically, in France, according to a Statista study of September 2019[67], of these approximately 1.7 million smart speakers, the estimated market share would be as follows (for the most popular brands): 49% Amazon Echo, 38% Google Home and 10% Siri. Add to this the JBL speakers including Google Assistant (8%) and Sonos, which carry Alexa (6%). However, these figures should be interpreted with caution as the CSA-Hadopi survey shows a larger share of Google's speakers compared to Amazon's, but which on the whole illustrate the domination of these two players (see Figure 3).

Figure 3

Distribution of smart speakers equipped with voice assistants in France in 2019

(source Statista).



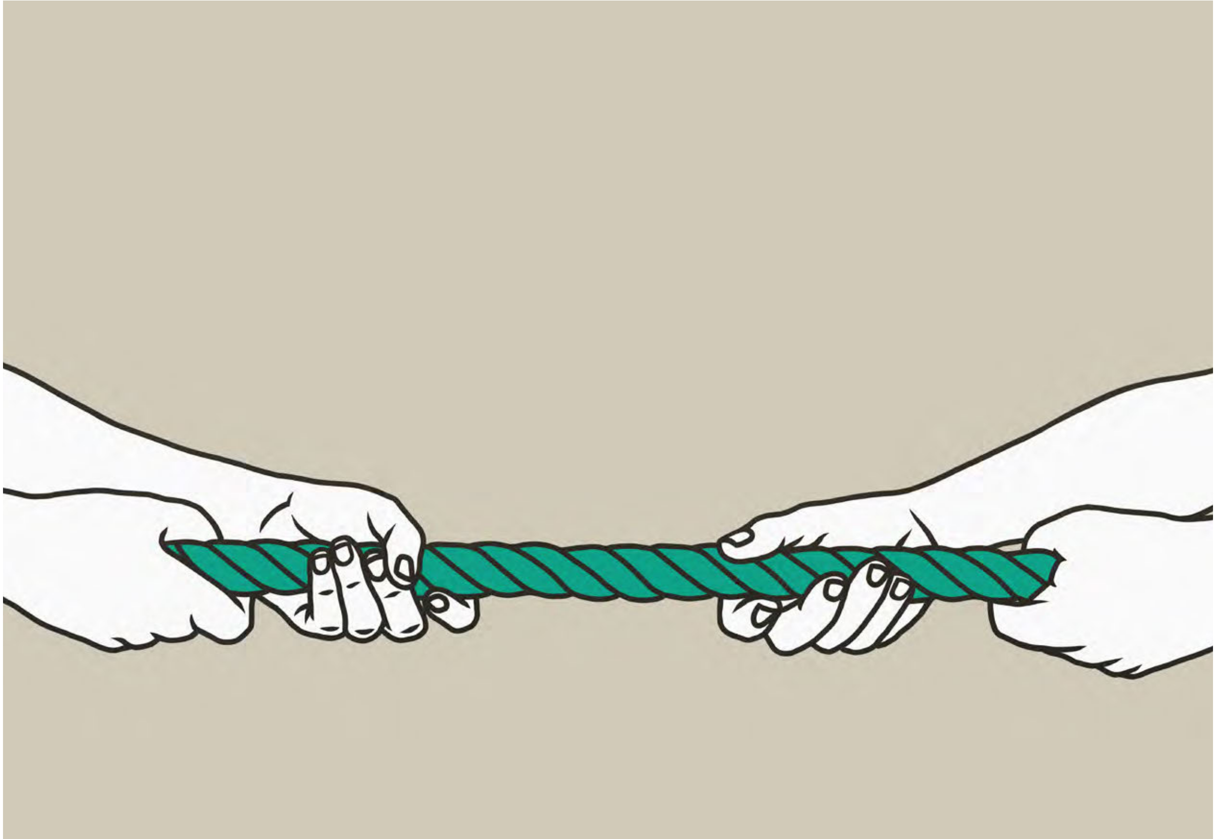
[63] - Dieter Bohn, Amazon says 100 million Alexa devices have been sold — what's next?, The Verge, january 2019, https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp

[64] - Scott Huffman, *Here's how the Google Assistant became more helpful in 2018*, Google keyword, january 2019, https://www.blog.google/products/assistant/heres-how-google-assistant-became-more-helpful-2018/

[65] - Jean-Sébastien Zanchi, *Aloha : le projet secret d'assistant vocal de Facebook*, 01net, august 2018, https://www.01net.com/actualites/aloha-le-projet-secret-d-assistant-vocal-de-facebook-1509984.html

[66] - Lucas Mediavilla, *Facebook développe son propre assistant vocal*, Les Echos, april 2019, https://www.lesechos.fr/tech-medias/hightech/facebook-developpe-son-propre-assistant-vocal-1012806

[67] - Statista, *Enceintes connectées : Amazon domine le marché français*, september 2019, https://fr.statista.com/infographie/19469/enceintes-connectees-les-plus-populaires-en-france/



TAPPING THE VOICE: MYTHS AND ISSUES ABOUT VOICE ASSISTANTS

After analysing the operation, practices and environment of voice assistants, it is now time to assess the consequences of their use. Design decisions, technology choices and marketing plans pose critical questions for users: the "voice object", by nature volatile and impalpable, finds rigidity and consistency through the use of these devices.

MYTHS AND REALITIES OF VOICE ASSISTANTS

Many myths circulate about voice assistants and the abilities that they are assumed to have. Let's have a closer look at five of them for a clearer understanding of the logic behind these systems and the questions they raise for their users.

MYTH #1: They're always listening

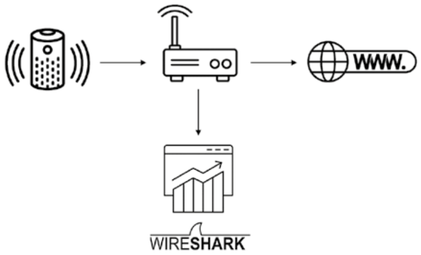
TRUE AND FALSE (they don't record everything!)

As specified in Chapter I.2 *Voice assistant, who are you?* and in particular in the infographics showing the generic operation of a voice assistant (page 14), in order to be used a voice assistant needs to be "awake". This means that the assistant switches to an active listening mode in order to receive orders and commands from its user. While in rare cases this "wakeup call" can be done by physical action (e.g. pressing a button, turning on the speaker, etc.), almost all voice assistants on the market rely on "keyword spotting" to switch to active listening mode (also known as wake word or hot word).

To do this, the assistant relies on the use of the microphone and slight computational capabilities to detect whether the wake word has been spoken. This analysis, which takes place continuously from the moment the assistant is on duty, is all carried out locally and only when the wake word has been recognised are the audio recordings processed

FOCUS ON...

LINC Experiments



Since 2017, the CNIL's Digital Innovation Laboratory (LINC) – the structure dedicated to the experimentation and study of emerging trends in digital usage – has been working on the subject of voice assistants. To this end, various tests and experiments have been carried out on the most widespread equipment on the market.

One of the tests carried out aimed in particular at observing the upbound network traffic coming from a voice assistant (in this case, the two most widely used assistants in France: Amazon Alexa and Google Assistant). For this purpose, a packet analyser (Wireshark) was used, placed at the cut-off point of the Internet connection (in order to observe the characteristics of the traffic flows between the assistant and the remote servers) as shown in the diagram above. In addition, a sound environment close to that produced by a household was simulated: the assistant is placed in a room comparable in size to a living room, conversations take place and a television is likely to be switched on at times. This installation demonstrated that, even though during the period of analysis no order specifically intended for the assistant had been formulated, many activations had been recorded and were visible in the usage history. Researchers at Northeastern University and Imperial College London have conducted similar experiments and obtained comparable results[68].

[68] - Daniel J. Dubois et al., *When speakers are all ears*, Smart speaker study, february 2020, https://moniotrlab.ccis.neu.edu/smart-speakers-study/

for interpretation and execution of the command, which in many cases results in transfer to remote servers over the Internet.

Wake word detection is based on machine learning techniques. The major challenge in using such methods is that the detection is probabilistic. Thus, for each word spoken, the system provides a confidence score as to whether the keyword was actually spoken. If this score turns out to be higher than a pre-set threshold value, it is considered that the wake word was indeed spoken. Such a system is therefore not error-free: in some cases, activation may not be detected even though the wake word has been uttered (false rejection) and in other cases activation may be detected even though the user has not uttered the keyword (false acceptance).

In practice, an acceptable compromise has to be found between these two types of errors to define the threshold value. However, since the consequence of a false wake word detection is the transmission of audio recordings, unexpected and unwanted feedbacks are likely to occur. Very often, developers of voice assistants implementing remote processing use a two-pass mechanism for this detection: a first one embedded locally at the equipment level and a second one carried out on the remote servers where the following data processing will take place (such as with Apple[69] or Google[70] for example). In this case, developers tend to set a relatively low threshold to enhance the user experience and ensure that when the user says the wake word, it is almost always recognised – even if it is "over-detected" – and then implement a second, more restrictive and computationally resource-intensive detection pass on the server side.

What this means is that, although in theory voice assistant providers have no intention of listening permanently, the intrinsically statistical nature of wake word detection makes the risk of false activation very real (see box above). There is therefore passive listening on the part of the voice assistants (apart from the activation button that has been set), which in some cases can have adverse consequences for users.

A question to…
Julia Velkovska and Moustafa Zouinar

These devices hold the promise of seamless interaction with the user. What's the situation in practice according to your observations?

First of all, automatic speech recognition, i.e. the transcription of human speech into text that can be used by the system, is not always efficient, even for simple requests such as asking for the weather forecast. Users sometimes have to repeat their statements several times to make themselves understood, which can lead them in some cases to stop using the system. In cases where they persevere, they engage in full-on "user work", linked to managing the interaction and its meaning. This work can result in a variety of actions such as rephrasing statements by shortening or expanding them to make them more precise, moving closer to the device, or speaking louder. This effort extends beyond the formulation of statements to encompass the full range of activities performed by individuals to make the system work, including those aimed at making sense of irrelevant responses or the learning work related to the imposed interactional "structure" (activating the system and then talking at the right time, i.e., when it is "listening"). This user work is a major aspect of the current uses of voice assistants and it is very important to describe it, understand it and take it into account when we think about the social consequences of the dissemination of these technologies at a time when they are coming out of the laboratory and taking their place in the heart of the home

Julia Velkovska and Moustafa Zouinar work as a sociologist and ergonomist respectively at the SENSE Laboratory, Orange Labs

> Full interview to be found on LINC [in French]
Julia Velkovska and Moustafa Zouinar : *Assistants vocaux : un véritable fossé entre les discours promotionnels et la réalité des usages*, Linc.cnil.fr, april 2018, https://linc.cnil.fr/fr/julia-velkovska-et-moustafa-zouinar-assistants-vocaux-un-veritable-fosse-entre-les-discours

[69] - Hey Siri: *An On-device DNN-powered Voice Trigger for Apple's Personal Assistant*, Apple Machine Learning Journal, october 2017, https://machinelearning.apple.com/2017/10/01/hey-siri.html

[70] - Assaf Hurwitz et al., *Keyword Spotting for Google Assistant Using Contextual Speech Recognition*, IEEE Automatic Speech Recognition and Understanding Workshop, december 2017, https://research.google/pubs/pub46554/

MYTH #2: They understand us perfectly

FALSE

While they implement multiple artificial intelligence (AI) techniques, ascribing the ability to produce intelligent behaviour, model abstract ideas, but also consciousness and feelings to a voice assistant still falls within the realm of science fiction. While some of these so-called "mainstream" assistants are likely to be adapted to respond to any type of request, they cannot be described as "strong AI".

Voice assistants are an avatar of what is more commonly known as "weak AI", i.e. increasingly autonomous systems implementing algorithms capable of solving problems of a given type. In contrast to the notion of "strong AI", the machine simulates intelligence and seems to act as if it were intelligent, for example by imitating the behaviour of one person in front of another during a dialogue.

Although, today, voice assistants provide satisfactory results on the whole, it should be noted that we are still far from the seamless, natural exchanges extolled by the promoters of these technologies. As Julia Velkovska and Moustafa Zouinar (see box) remind us, it is thus necessary for the user to genuinely "work". Indeed, when two people speak together, we observe a phenomenon of interactional convergence during which each person takes on the discursive habits of the other in order to create a common space that will promote mutual understanding. In the case of an exchange with a voice assistant (or any other talking machine), it is the user who, alone, operates this convergence in order to adapt to the comprehension capacities of the machine.

Finally, the reason why users are sometimes shocked by the slip-ups and failures of voice assistants may be due to the "uncanny valley" theory invented in the 1970s by Japanese robotic engineer Masahiro Mori. According to this theory, the more similar a robot is to a human being, the more monstrous the differences between them appear. Transposed to the case of voice assistants, this implies that the more questions an assistant is able to answer, the more the user is struck when it fails. It is therefore necessary to precisely define an assistant's field of intervention at the time of its design, otherwise it will lead to real disillusionment among its users.

MYTH #3: They use our data to better profile us

TRUE AND FALSE

We saw in Chapter I.4 *What strategies for voice assistant designers?* that the designers of voice assistants adopt different business models. In some of them, the aim is to provide a service to a user with the sole purpose of allowing the user a new way of interacting with equipment using voice. This is particularly the case with the white label embedded voice assistants requested by some TV manufacturers, vacuum cleaners, etc. In these cases, the user's interactions with his/her assistant do not generally feed into an ecosystem of data.

However, for the main providers of assistants, in particular the leading tech players the likes of Google and Amazon, whose business relies on the processing of their users' personal data, this is only a new vector of collection. Primarily intended for use in the home to control connected objects, entertainment services or home automation applications, devices equipped with a voice assistant are at the heart of home life. Often, it is necessary to create a user account to take full advantage of the options offered by the assistant. The major manufacturers thus propose to link exchanges with their assistants directly to already existing user accounts to use their products (email, calendar, online store, etc.). Thus, the user profile is built (in the case of a new account) and added to (in the case of an existing account) by the user's different interactions with the assistant: lifestyle habits (wakeup and bedtimes), heating settings, cultural tastes, past purchases, interests, etc. By enhancing the information already existing about a user in this way, it is possible to implement more targeted advertising campaigns and to offer more detailed commercial proposals correlated to the information collected, etc. However, such a positioning reveals a tension between the implementation of a profiling model centred on the individual and the potentially very collective nature of the assistant, for example when deployed in a household. While the voice assistant can collect a lot of information about the group in which it is deployed, a multiplication of users can also be confusing for personalised models, with for example commercial proposals or advertisements for goods or services that could be made as a result of other people's interactions with the assistant (spouse, children, etc.).

For players implementing such profile-feeding strategies, it is essential to promote their role as intermediary and orchestrator to third-party professionals (connected object providers, online vendors, transport organisations, etc.) in order to convince them of the merits of being present on their platform by developing an ad-hoc application. Indeed, the greater the offering on these platforms, the more frequently users are likely to interact with the assistant and the more detailed the information held about them. Therefore, the number of addressable devices, i.e. equipped with the assistant, is a decisive sales argument to encourage third-party application developers!

On a final note, although, for the time being, profiling is carried out solely on the basis of information inferred from orders placed, the question of broadening the scope of the collection may arise, without prejudging the lawfulness of such practices. For example, information about ambient noise obtained by analysing sound scenes (TV channels watched, young child crying in the background, etc.) or emotional or health status obtained by analysing the voice signal may be of interest to assistant developers. For example, in 2017 Amazon filed a patent to categorise the emotional states of its users[71] and launched in August 2020 Halo, a fitness band and app analysing the body and voice.

MYTH #4:
They're a popular interface for children

TRUE

As seen in the previous chapter, these devices appeal strongly to children. They allow them – at least in theory – to access resources that are usually beyond their reach, such as a hi-fi system, and thus to play the music of their choice. However, children's convoluted and faltering utterances are often mistranscribed and therefore also misinterpreted, causing them significant frustration and also forcing them to "work" to adapt to the assistant's abilities (see Myth 2).

In 2014, Johan Schalkwyk, head of speech processing at Google, announced that "our children and grandchildren won't understand that there was once a time we were able to interact with a computer other than by voice". Very early on, the tech giants identified voice as the human-machine interface (HMI) of the future, easily accessible by everyone and everywhere. The leading players' advertising campaigns also focus on this aspect, as the assistant is supposed to allow easy, seamless interaction, including for audiences least familiar with digital codes and practices, i.e. young children and the elderly. The communication of Facebook on its Portal tool illustrates this well: it's about keeping the whole family in touch, from children to grandparents. Young audiences are also offered content specifically produced for them: applications of jokes, games, stories that are read-out, etc., many of which actually conceal a commercial purpose since they are developed by this or that brand. What's more, the "education" of future consumers is also being shaped by the manipulation of this interface and grammar of use.

While voice assistants may offer an alternative to screens, there are questions about leaving such a device freely accessible to children. There are many examples that exist to illustrate the potential risks: shopping online[72], access to adult content[73], etc.

In general, and as with all connected objects (including toys), it is essential to implement basic good protection practices. The CNIL had the opportunity to communicate on this subject[74]. Applied to the case of voice assistants, it is particularly important to ensure the use of parental filtering techniques and to supervise children's interactions. For example, Mattel abandoned the idea of launching the Aristotle voice assistant that was to equip its products, offering a platform on which parents could listen to and even play back the conversations that children had with their toys[75]. A practice that can potentially harm the child's privacy or affect the relationship of trust between the child and his or her parents.

[71] - Jon Brodkin, *Amazon patents Alexa tech to tell if you're sick, depressed and sell you meds*, Ars Technica, november 2018, https://arstechnica.com/gadgets/2018/10/amazon-patents-alexa-tech-to-tell-if-youre-sick-depressed-and-sell-you-meds/

[72] - Karma Allen, *6-Year-Old Mistakenly Orders Dollhouse, Cookies Worth $162 While Chatting With Amazon Echo*, ABC News, january 2017, https://abcnews.go.com/Technology/year-mistakenly-orders-162-worth-treatschatting-amazon/story?id=44577327

[73] - Post staff report, *Boy requests song from Amazon Alexa, but gets porn instead*, New York Post, december 2016, https://nypost.com/2016/12/30/toddler-asks-amazons-alexa-to-play-song-but-gets-porn-instead/

[74] - CNIL, *Jouets connectés : quels conseils pour les sécuriser ?*, https://www.cnil.fr/fr/jouets-connectes-quels-conseils-pour-les-securiser

[75] - Rachel Rabkin Peachman, *Mattel Pulls Aristotle Children's Device After Privacy Concerns*, The New York Times, october 2017, https://www.nytimes.com/2017/10/05/well/family/mattel-aristotle-privacy.html

MYTH #5: They're hackable

TRUE

Embodied in smooth, sleek equipment, it is common to forget the purely software nature of voice assistants. And yet, these belong by nature to the galaxy of the Internet of Things (IoT). As such, they are equally subject to computer attacks.

The number of attacks actually recorded against voice assistants is still quite low today. Known and reported attacks in recent years appear to have been mainly carried out by security researchers (see box). However, as with all Internet of Things equipment, it is to be expected that, with the development of the services offered, hackers will find increasing merits in accessing these devices illegitimately, either to take control of them or to access the data passing through them. What's more, with the ambition to make voice assistants the nerve centre of the connected home (smart hub), they can become a central point of vulnerability for the "smart" home's information system.

FOCUS ON...

Overview of breaches and attacks

The first perpetuated (and documented) attacks against voice assistants seem to have been those implemented by the Burger King company[76] and the South Park series[77] in 2017. In both cases, audiovisual content containing the wake words activating the main assistants ("Hey Alexa", "Ok Google") followed by an order to transmit advertising or far-fetched information was broadcasted. In these two examples, the attackers rely on two facts: 1) a voice assistant can be activated by anyone at a listening distance, and 2) smart speakers are often placed in the living room, right where the home TV is located. This attack, although rudimentary, has forced designers of voice assistants to deploy bypass strategies that prevent speaker activation either by performing an audio signature of the content in order to identify and block it (which implies prior exposure), or by performing continuous analysis of voice assistant activations and blocking the execution of the command when too many simultaneous activations are observed.

Since then, the cybersecurity research community has largely taken up this new subject of study of voice assistants. A large number of attacks have shown that the most famous voice assistants can be activated and used without their owners' knowledge. The first major type of attacks, like the Dolphin attack[78] (but also the Backdoor attack[79] and LipRead attack[80]) is based on the fact that voice assistant microphones are sensitive to high frequencies outside the audible spectrum (above 20,000 Hertz). This makes it possible, using rudimentary equipment (smartphone, amplifier and ultrasound transducer) to pass imperceptible commands onto an assistant. The second family of attacks uses the psychoacoustic phenomenon of "masking" to hide commands to voice assistants in other audio signals (such as the chirping birds in the Chirping birds attack[81]).

(continued P36...)

[76] - Antoine Boudet, *Burger King détourne Google Home pour faire sa publicité, Google riposte*, Numérama, april 2017, https://www.numerama.com/tech/249062-burger-king-detourne-google-home-pour-faire-sa-publicite-google-riposte.html

[77] - Vincent Tanguy, *Quand South Park rend fous Alexa et Google Home*, Sciences et Avenir, september 2017, https://www.sciencesetavenir.fr/high-tech/intelligence-artificielle/quand-south-park-rend-fous-alexa-et-google-home_116434

[78] - Guoming Zhang et al., *DolphinAttack: Inaudible Voice Commands*, ACM Conference on Computer and Communications Security, november 2017, https://acmccs.github.io/papers/p103-zhangAemb.pdf

[79] - Nirupam Roy et al., *BackDoor: Making Microphones Hear Inaudible Sounds*, MobiSys 2017, june 2017, https://synrg.csl.illinois.edu/papers/backdoor_mobisys17.pdf

[80] - Nirupam Roy et al., *Inaudible Voice Commands: The Long-Range Attack and Defense*, USENIX Symposium on Networked Systems Design and Implementation, april 2018, https://synrg.csl.illinois.edu/papers/lipread_nsdi18.pdf

[81] - Lea Schönherr et al., *Adversarial Attacks Against ASR Systems via Psychoacoustic Hiding*, Network and Distributed System Security Symposium, february 2020, https://adversarial-attacks.net/

In a more general way, we are talking here about adversary attacks which allow, as demonstrated in several studies[82][83], to hide any command in an audio recording that cannot be detected without computer analysis. Other, even more surprising, attacks have also been carried out. The Surfing attack[84] allows ultrasonic commands to be injected by vibrating the surface on which the voice equipment-enabled device is placed (e.g. a table). Finally, by a laser attack[85], a hacker can send inaudible and invisible commands by pointing a laser at the speaker's cone, and the attack can be carried out from a distance of up to 100 metres!

Several research papers have also shown that voice assistants can be used for "vishing" (voice phishing) purposes. The technique of Skill squatting[86], allows a hacker to create a homophone application of a legitimate application based on the limitations of the speech recognition system. For example, researchers are developing an application called "Am Express" that can be used to divert requests made to American Express' "Amex" application and interact with a user to access their personal and financial information. Other illegal phishing and eavesdropping techniques are also used[87]. These particularly rely on making the user believe that the interaction with the application s/he has called is over, when it is actually still running, for example, by asking the assistant to play an unpronounceable sequence that silences the assistant, but leaves the application running.

It should also be remembered that the specific characteristics of the assistant make it vulnerable without requiring a high level of technical knowledge. Indeed, the activation radius of an assistant is around five metres: being within this range can allow an ill-intentioned person to use the assistant (open a door, retrieve information, etc.). In the same way, the assistant works on a sensor principle (a microphone), which can be scrambled: by creating ambient noise, for example, whether or not it is audible to humans (white noise that covers all frequencies uniformly). Therefore, it would be possible to prevent the operation of certain devices relying on speech recognition alone. Some, like the Alias Project[88], incidentally use this principle to prevent the untimely activation of voice assistants. This involves a protective overlay that is placed over a loudspeaker equipped with a voice assistant. It permanently diffuses white noise except when a wake word, defined beforehand by the user, is spoken. From that moment on, the emission of the extraneous noise stops and the voice assistant becomes usable in the traditional way.

In addition to attacks, whether cyber-related or otherwise, failures are also likely to occur and several of them have already been made public. The Google Home mini devices – one of the special features of which was that it could be activated simply by tapping it – came under scrutiny in 2017 when a journalist noticed that his device was activated thousands of times a day when he checked his usage history[89]. The device detected "ghost touches" and was therefore permanently activated. Finally, in another notable misstep, during a subject access request made to Amazon, a German user obtained a history of interactions with Alexa from another Amazon account holder. Investigative work by journalists led to the identification of the person to whom the data belonged, who subsequently filed a complaint[90].

[82] - Tavish Vaidya et al., *Cocaine Noodles: Exploiting the Gap between Human and Machine Speech Recognition*, USENIX Workshop on Offensive Technologies, august 2015, https://www.usenix.org/node/191969

[83] - Nicholas Carlini et David Wagner, *Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, IEEE Security and Privacy Workshops (SPW), mai 2018, https://arxiv.org/pdf/1801.01944.pdf

[84] - Qiben Yan et al., *SurfingAttack: Interactive Hidden Attack on Voice Assistants Using Ultrasonic Guided Waves*, Network and Distributed System Security Symposium, février 2020, https://surfingattack.github.io/

[85] - Takeshi Sugawara et al., *Light Commands: Laser-Based Audio Injection on Voice-Controllable Systems*, novembre 2019, https://lightcommands.com/

[86] - Deepak Kumar et al., *Skill Squatting Attacks on Amazon Alexa*, USENIX Security Symposium, août 2018, https://www.usenix.org/conference/usenixsecurity18/presentation/kumar

[87] - Security Research Labs, Smart Spies: *Alexa and Google Home expose users to vishing and eavesdropping*, novembrenovembre 2019, https://srlabs.de/bites/smart-spies/

[88] - Bjørn Karmann, *Project Alias*, 2018, http://bjoernkarmann.dk/project_alias

[89] - Artem Russakovskii, *Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7*, Android Police, octobre 2017, https://www.androidpolice.com/2017/10/10/google-nerfing-home-minis-mine-spied-everything-said-247/

[90] - Holger Bleich, *Amazon reveals private Alexa voice data files,* Heise online, december 2018, https://www.heise.de/newsticker/meldung/Amazon-reveals-private-voice-data-files-4256015.html

WHAT ISSUES FOR VOICE ASSISTANTS?

Ethical issues...

The appearance of these new interaction devices on smartphones as well as in the home raises ethical questions. Since they change the way we search for and find information, live in our home (and within it via home automation), interact with personal services (email, calendar or even bank account), they affect our environment in the broadest sense. It is therefore relevant to consider these changes and their consequences both in terms of operation and uses.

FOCUS ON...

The CNPEN and the ethical issues of conversational agents

The National Pilot Committee for Digital Ethics (CNPEN) was established in December 2019 at the request of the Prime Minister[91]. Made up of 27 members, including a representative of the CNIL, this committee brings together digital specialists, philosophers, doctors, lawyers and members of civil society. One of the three referrals submitted by the Prime Minister to the CNPEN concerns the ethical issues of conversational agents, including in particular voice assistants. In this context, and to prepare its recommendations for the attention of designers as well as users of conversational agents, the CNPEN launched a call for contributions[92] intended to give stakeholders and the public an opportunity to express their views on the ethical issues related to these chatbots that are increasingly present in our lives. The ethical issues presented here chime with the CNPEN's ethical considerations, which build on the work initiated by CERNA, the Allistene Alliance's Commission for Reflection on the Ethics of Research in Digital Science and Technology.

1) From screen to voice, the creation of a new grammar of use

Voice assistants propose a major paradigm shift by transforming the relationship with digital tools, through the transition from visual to vocal. The graphical interfaces are gradually fading away, and these devices, with seamless interaction in mind, leave the door wide open for voice. The screen remains a secondary accessory, through "companion" accounts and applications, which are particularly used to configure the system and monitor its activity. However, the goal is to do away as much as possible with any visual accessory or visible buttons. As we have seen above, this new human-machine interface offers up huge advantages (see Chapter I.3 *What can voice assistants be used for?*), but it also raises questions. Without a screen, it is difficult to have an overview of the traces recorded, to judge the relevance of the suggestions, to find out more or to trace the information that is provided. At a time when many tech players are seeking to make “technology disappear”, there is a real risk of these systems becoming opaque and the methods of making information accessible and obtaining consent must be carefully thought out. Can we be well informed vocally? As the first place of interactions between the individual and information and between the individual and his/her rights, the interfaces require a solid and clear design allowing everyone to understand the stakes of using a service or product. For several years now, the CNIL has been working on the design of these systems, leading to the publication in January 2019 of the 6th Innovation and Foresight Report “Shaping Choices in the Digital World”[93], as well as a dedicated Data & Design website[94].

[91] - National Consultative Ethics Committee, *Création du Comité Pilote d'Éthique du Numérique*, december 2019, https://www.ccne-ethique.fr/fr/actualites/creation-du-comite-pilote-dethique-du-numerique

[92] - CNPEN, *Les enjeux éthiques des agents conversationnels*, june 2020, https://www.ccne-ethique.fr/fr/actualites/cnpen-les-enjeux-ethiques-des-agents-conversationnels

[93] - LINC, *IP6 Report: Shaping Choices in the Digital World*, january 2019, https://linc.cnil.fr/fr/ip-report-shaping-choices-digital-world

[94] - CNIL, *Données & Design : une nouvelle plateforme pour la communauté des designers autour du RGPD*, https://www.cnil.fr/fr/donnees-design-une-nouvelle-plateforme-pour-la-communaute-des-designers-autour-du-rgpd

2) From several pages of results to a single answer

The relationship with the voice assistant also brings about a paradigm shift in the way information is sought online. Contrary to the conventional operation of a search engine, the assistant will choose a single answer to the question asked and deliver it as is. There will be a single answer to a question, where previously it was possible to navigate through the results pages. This is therefore a shift from a search engine to a response engine. Providing only one answer raises questions about the choice of result sources. For simple questions, at least in appearance, the problem may not arise (weather, encyclopaedia, etc.). In other cases, when it comes to political news, for example, the choice of sources to be given precedence arises. How can their neutrality be ensured? And how do you ensure their reliability? Do the answers vary according to the profile of the individual? Is this a good thing?

The example of Wikipedia is particularly telling. The content proposed by the editors is corrected, verified and moderated by the community, generating a neutral effect. But because of this openness, Wikipedia can temporarily fall victim to misappropriation. Changing content can then contribute to the dissemination of biased or advertising statements, such as those that The North Face published to promote its products in 2019 (which earned it scathing criticism, forcing it to issue an apology[95]). Cases of misappropriation containing aggressive language have also been documented. For example, a version of a Wikipedia article on the cardiac cycle that advised "stabbing yourself in the heart" because heart activity contributed to "global impoverishment" was read by voice assistants[96]. In this case, it was a hoax by an ill-intentioned Wikipedia contributor, whose changes were quickly scrapped by the community.

3) From the human-machine interface to the human-android relationship?

a) A tendency towards anthropomorphisation

Synthetic voices have not waited for the development of voice assistants to break into the lives of individuals and populations, particularly in urban areas (in stations, public transport, at pedestrian crossings, etc.). In order to achieve greater acceptance on the part of users, the designers opted for personification, preferred over an artificial and mechanical model. In the specific case of voice assistants, this choice was made with a view to greater uptake. By reproducing the conditions of a dialogue in natural language, interactions with the tool are made easier and thus more numerous. Individuals can ask their questions directly to the assistant in the same way as they would ask another person, with the addition of the wake word. For its part, the assistant takes the time to answer according to established language codes (subject - verb - complement), and not only by delivering unprocessed information. By giving it a humanised name (Siri, Alexa, Cortana, Célia, etc.), characteristics and a story, by making it tell jokes, through the richness of its answers, the designers have attached a fiction and a personality to their voice assistant. The goal is to go beyond a simple tool to become a partner and companion. Moreover, the user is prompted to configure his/her device, in particular as regards the characteristics of the voice with which s/he will interact (gender, accents, etc.). Information and communication science researcher Clotilde Chevet clarifies this aspect in her work by returning to the concept of the Ikea effect. The latter, developed by researchers Shyam Sundar and Yuan Sun, shows a stronger attachment to an object if we contributed to its constructione[97].

b) Gender characteristics of the assistant

The characteristics of voice assistants (names, voices) can contribute to the reproduction of an unequal representation of women's place in society, starting with placing them in the role of the assistant. In 2019, UNESCO published a report on this issue, entitled "I'd blush if I could", referring to what Siri would say when the voice assistant was misogynously insulted[98]. The assistant's response had sparked much criticism and changes were subsequently made: "I don't know how to answer that". The observation of the reproduction of gender stereotypes through voice interfaces has led to the development of projects for gender-neutral voice assistants such as Q[99].

c) Confusion between virtual agent and human interlocutor

The relationship with these objects also raises a number of questions. How should these new devices be considered within the home space? What importance should we attach to them in our homes? And within the family? As seen above, younger people account for a large proportion of users of voice assistant-enabled smart speakers, which raises questions in terms of education in particular. Titiou Lecoq, journalist and member of the CNIL Foresight

[95] - LeBrief, *The North Face détourne Wikipédia pour placer ses produits, puis s'excuse*, NextInpact, may 2019, https://www.nextinpact.com/brief/-the-north-face-detourne-wikipedia-pour-placer-ses-produits--puis-s-excuse---8860.htm

[96] - James Crowley, *Woman Says Amazon's Alexa Told Her To Stab Herself In The Heart For 'The Greater Good'*, Newsweek, december 2019 https://www.newsweek.com/amazon-echo-tells-uk-woman-stab-herself-1479074

[97] - Clotilde Chevet, *La voix de synthèse : de la communication de masse à l'interaction homme-machine.* Dialogue avec le monde, Communication & langages 2017/3 (N° 193), https://www.cairn.info/revue-communication-et-langages1-2017-3-page-63.htm

[98] - UNESCO, EQUALS Skills *Coalition, I'd blush if I could: closing gender divides in digital skills through education*, 2019, https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

[99] - https://www.genderlessvoice.com/

Committee, has highlighted the questions that arise in the relationship between a child and the assistant[100]. Indeed, should we say thank you to them? How do you make a child understand the difference between a real (human) person and artificial intelligence (AI)? How can you resist the possibility of giving an order to an adult voice that will never be able to tell you off?

4) Voice assistants are AI systems like any other

a) The presence of bias

Voice assistants use artificial intelligence technology. They are grounded in machine learning methods for a wide range of tasks: wake word detection, automatic speech transcription, language comprehension, speech synthesis, etc. In 2017, the CNIL published an ethical report entitled "How can humans keep the upper hand? "[101] that questions the ethical issues of algorithms. The question of the biases that learning systems are likely to harbour is discussed in this report, as well as the risks that they imply for users. Indeed, these computer systems learn from data sets that have been collected, selected, labelled, etc., but which may contain biases. Over-representation or under-representation of certain populations or characteristics can influence the learning of AI and subsequently pass this error or misconfiguration on to its calculations, and therefore in its way of functioning[102]. Data quality therefore plays a major role in the finesse and accuracy of learning, just as much as data quantity. Where voice assistants are concerned, the text corpora used can statistically bring up female pronouns more often than male ones, and the voices used for training may be mainly those of adults, whereas the system is designed to be able to interact with children, etc. Moreover, different voice timbres, ways of speaking, accents, languages, etc. can introduce specific sources of error in speech recognition. While it is possible to anticipate these biases and inject patches into systems, not all biases are necessarily known in advance.

b) The use of clickworkers

Whether it is to qualify the learning database or to correct errors made when the algorithm is deployed, learning and training of artificial intelligence systems necessarily requires human intervention. This part of the work, known as "digital labour", has been criticised. In his work, sociologist Antonio Casilli describes the outsourcing of these micro-tasks to low-skilled and low-paid people, sometimes in low-wage areas of the world, in order to reduce costs[103]. This raises questions about both working conditions and safety. In many cases, it has been observed that data, the raw material for these micro-tasks, can circulate between designers of artificial intelligence systems and subcontractors without the necessary guarantees being put in place[104].

...and more specifically, privacy issues

Interactions with assistants feed into a collection of information about daily routines and individual's privacy. Moreover, this collection is amplified by the use of third-party applications, those services used through voice assistants and which allow you to consult your bank account balance or make transfers, consult your statements, control the roller shutters or lights in your home, monitor your energy consumption, etc. The increasing exchange of data and information stored on different accounts linked to individuals raises questions about privacy and data protection.

1) Private, potentially sensitive data

As discussed in Chapter I.1 *The specificity of the voice*, the voice can contain a great deal of information about the words spoken by the individual and his or her identity or inferred characteristics such as emotional state, socio-cultural background, ethnicity or health. The voice reveals information that concerns our private lives. Where such information reveals racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership or concerns state of health, sex life or sexual orientation, it falls within the category of sensitive data, specially protected by data protection laws (Article 9 of the GDPR). Voice contains markers specific to an individual, a combination of physiological and behavioural factors. This is what makes it a biometric attribute in its own right, which can be used to identify the individual. Voice assistants can offer to create profiles linked to an individual's voice, so that several accounts can be associated with the same voice assistant (for example, different members of a household). In practice, this means identifying the speaker among the profiles enrolled simultaneously with the pronunciation of the wake word used to activate the assistant. The devices can thus associate several people and not confuse them when one of them requests access to personal information (emails, calendar, etc.). However, voice recognition is subject to a number of constraints: microphone quality, ambient noise, physical condition of the person, etc. that may hinder this identification. Finally, the circulation of biometric data, or data enabling the creation of biometric templates, and

[100] - Titiou Lecoq, *Les enfants pensent que les enceintes connectées sont vivantes et c'est un problème*, Slate.fr, june 2018 http://www.slate.fr/story/163907 enceintes-connectees-intelligence-artificielle-alexa-google-home-siri-education-enfants

[101] - CNIL, How can humans keep the upper hand? Report on the ethical matters raised by algorithms and artificial intelligence, december 2017, https://www.cnil.fr/en/ how-can-humans-keep-upper-hand-report-ethical-matters-raised-algorithms-and-artificial-intelligence

[102] - Koenecke et al., *Racial disparities in automated speech recognition, Proceedings of the National Academy of Sciences*, avril 2020, https://www.pnas.org/content/117/14/7684

[103] - Antonio A. Casilli, *En attendant les robots : Enquête sur le travail du clic,* Seuil, 2019

[104] - Alex Hern, *Skype audio graded by workers in China with 'no security measures'*, The Guardian, janary 2020, https://www.theguardian.com/technology/2020/jan/10/ skype-audio-graded-by-workers-in-china-with-no-security-measures

the conditions for storing them – potentially carried out remotely – raise questions as to their possible retrieval for hacking purposes. If clandestine spaces, such as the dark web, already abound with databases containing email addresses, passwords and other information[105], in the future, an intense use of voice interfaces could make these biometric features a highly traded commodity on data black markets.

2) Recording devices increasingly present in shared spaces

a) From home to collective and work places

Voice assistants, initially personal and associated with a smartphone, have liberated themselves from this individual notion to be deployed in new spaces. First present in the home through smart speakers (living room, bedroom, kitchen, etc.), they can now be found in places where people pass by. For example, they are increasingly being installed in hotels, to offer new services to customers, or in workplaces, such as doctors' surgeries, for example, to make appointments easier. This deployment of voice assistants in these transit areas raises questions about the confidentiality of the exchanges that are held, but also about professional secrecy (particularly in cases related to medical secrecy, in the context of the relationship with a lawyer, for the preservation of the secrecy of sources, etc.). To give an example, fears of a data leak or hacking of these listening devices led one of Ireland's largest law firms to ban them within the company, in particular for employees who work remotely and have one in their home[106]. Finally, the dissemination of these tools in professional settings also raises the issue of employee monitoring. The way in which they are used could enable permanent and constant monitoring[107].

b) What information should be given to people within the assistant's range?

Since voice assistants have a range of several metres (about five metres indoors), informing people in the vicinity is a recurring pitfall. While they have not configured the device themselves, their voices can be recorded, either through the activation of a voice assistant they do not own or through the unwanted capture of a conversation held near one of these recording devices. However, access to information and the exercise of rights must also be guaranteed for these people and several questions arise. How do you deal with the fact that third party data may be captured without their knowledge? How do you secure the primary user's information if third parties are likely to interact with the object? Paradoxically, one solution may be to use even more data to secure and personalise certain interactions with the assistant, for example by using speaker recognition.

c) Devices that blur the distinction between private and public spaces

The use of voice-based devices questions the separation between private and public space and redefines the notion of privacy. Speaking or receiving audio in a shared space can have a variety of consequences for individuals: increased exposure of one's privacy to third parties, risk of false manipulation and untimely activation, possibility of untimely activation by simply pronouncing the wake word, etc. These new forms of interaction in the social space are changing the way users relate to their private lives.

3) Devices that involve many players

The operation of voice assistants involves the intervention of a number of players and intermediaries throughout the implementation chain. Between the designers of voice assistants and the users are the integrators, third-party application developers and possibly the "deployers" of these assistants (see Chapter I.2 *Voice assistant, who are you?*). This raises the question of the responsibilities of each actor. Who is the data controller? How are the relationships between the assistant designer and the application developers organised? As illustrated in the infographics on page 14, data flow patterns are multiple and vary according to usage and design choices. Case-by-case analyses should therefore be carried out in order to clarify the different roles, modalities of intervention and capacities for action of each actor.

4) Human eavesdropping for product enhancement purposes

Following a number of high-profile scandals in the summer of 2019, all the major players in the voice assistant market (Amazon, Google, Microsoft, Facebook, and Apple) revealed that the audio recordings made by their assistants were, for some of them, listened to by individuals, either directly employed by these companies or acting as subcontractors, in order to categorise utterances, improve the quality of wake word detection, improve the performance of speech transcription and interpretation systems, and so on. Indeed, while the practice of human monitoring and annotation is indispensable for machine learning systems, these scandals underline the lack of

[105] - Lily Hay Newman, *1.2 Billion Records Found Exposed Online in a Single Server*, Wired, november 2019, https://www.wired.com/story/billion-records-exposed-online/

[106] - Aaron Rogan, *Law professionals banned from working at home near Alexa devices*, BusinessPost, february 2020, https://www.businesspost.ie/legal/law-professionals-banned-from-working-at-home-near-alexa-devices-3a681a75

[107] - CNIL, *Video surveillance - video protection at work*, https://www.cnil.fr/fr/la-videosurveillance-videoprotection-au-travail

clear information for data subjects regarding this replay of past (real or supposed) interactions with an assistant. In response, the Hamburg data protection authority ordered Google to suspend temporarily (for a period of three months) the activities relating to the listening and manual transcription of the words collected by its assistant[108]. Subsequently, all the major players in the voice assistant market announced that they had put an end to these practices or introduced an opt-out system for users[109][110]. At the same time, such listening-in for the purpose of improving the system has raised other questions. Some people in these eavesdropping-based professions have reported hearing recordings that were described as disturbing or even illegal, extending to assault in some cases[111][112]. Finally, the seizure by courts of recordings made by voice assistants during criminal investigations also raises questions about the legal implications that the use of voice assistants could have[113].

5) False positives and outsiders: the issue of consent and control over one's data

The GDPR aims at restoring the sovereignty of our data and our ability to exercise our rights. In the case of voice assistants, the actual operation of the device requires ex post control. The individual only has access to his or her activity and records (if retained) once they have been processed and, depending on the implementation, sent to remote processing servers. It is then possible to play back or delete the recordings, including some false positives, i.e. untimely triggering of the assistant following a false detection of the wake word. The simple fact of being able to know what was transmitted to the servers only by consulting one's account further begs the question of when this data was not meant to be transmitted. Loss of control of an account or its hacking can therefore lead to access not only to a consented activity (when using the device intentionally), but also to a more personal part, which may be more private and not necessarily known to the owner of the assistant. In addition, there may be people who are not aware of the recording device, but whose words are captured regardless. They may be kept without their knowledge and may also remain accessible to the owner.

6) Privacy by design and virtuous initiatives

Some voice assistant designers are choosing to develop products that respect privacy and personal data protection, such as the startup Snips (acquired by Sonos in 2019) or Mycroft, which has developed an open-source voice assistant solution[114].

Article 25 of the GDPR recalls the importance of user privacy protection by design and by default. Adapted to the case of voice assistants, it is possible to build an approach to the data cycle that is more respectful of the rights of individuals. Indeed, it seems relevant to categorise the types of use of the assistant and to associate models with these allowing the data transferred to be minimised – and, to a certain extent, greater restraint in the use of digital data. First, for internal uses (taking notes, setting a timer, etc.) or home automation (turning on the light, lowering the shutters, etc.), local processing of information may present less risk to the privacy of users than processing relying on the use of remote servers for the same result. Such a design choice requires embedding more "intelligence" (computational capacities) into the devices. From this point of view, relations and exchanges with elements external to the voice assistant-enabled equipment would be reduced to two main uses: 1) actions requiring access to remote knowledge bases (accessing online information, making an appointment, making a call, listening to music on a streaming site, etc.) and 2) feedback for the purposes of product enhancement or usage statistics.

Environmental impact, business model, dominance and competition: other issues for voice assistants

As with many other devices, the environmental impact of voice assistants raises questions, particularly in terms of their energy consumption. Whether in terms of production (through mass production of these new devices), data transit (through remote server-based operation) or the computing capacity required for voice processing, voice assistants are energy-intensive. Thus, as highlighted by the ANR DAPCODS/IOTics project (see box page 45), the traffic generated by a loudspeaker can be counted in hundreds of kilobytes, which is two orders

[108] - The Hamburg Commissioner for Data Protection and Freedom of Information, Speech assistance systems put to the test - Data protection authority opens administrative proceedings against Google, august 2019, https://datenschutz-hamburg.de/assets/pdf/2019-08-01_press-release-Google_Assistant.pdf

[109] - Guillaume Périssat, *Assistants vocaux : Apple et Google suspendent les écoutes tandis qu'Amazon propose de les désactiver*, L'informaticien, august 2019, https://www.linformaticien.com/actualites/id/52597/assistants-vocaux-apple-et-google-suspendent-les-ecoutes-tandis-qu-amazon-propose-de-les-desactiver.aspx

[110] - Lidia Davis, *How to opt out of Google Home's tracking features*, Reviews.com, march 2020, https://www.reviews.com/home/smart-home/opt-out-of-google-homes-tracking-features/

[111] - Matt Day, Giles Turner, and Natalia Drozdiak, *Amazon Workers Are Listening to What You Tell Alexa*, Bloomberg, april 2019, https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio

[112] - Anouk Helft, *Apple admet écouter certaines conversations privées via Siri*, Les Echos, july 2019 https://www.lesechos.fr/tech-medias/hightech/apple-admet-ecouter-certaines-conversations-privees-via-siri-1041502

[113] - LePoint.fr, *L'enceinte Alexa a-t-elle été témoin d'un meurtre ?*, Le Point, november 2019 https://www.lepoint.fr/monde/l-enceinte-alexa-a-t-elle-ete-temoin-d-un-meurtre-07-11-2019-2345984_24.php

[114] - https://mycroft.ai/

of magnitude more than for connected objects that do not have automatic speech processing capabilities. So, asking your voice assistant to turn on a light does not have the same impact as using a switch – and this difference is not necessarily noticeable to the user. What's more, the natural language models used in voice assistants have improved considerably at the cost of a significant increase in energy costs. As an example, Google's training of the BERT model (see box page 19) required the learning of some 340 million parameters for an electricity cost equivalent to an American household's consumption for 50 days. Today, some language models such as OpenAI's GPT-3 or Nvidia's MegatronLM have billions of parameters[115]! These examples clearly illustrate the cost of digital technology, which it is now essential to take into account, including in the field of personal data protection. To some extent, the principle of data minimisation echoes the concept of restraint in the use of digital data and also aims to create systems that are optimized in data use. The CNIL has also initiated discussions on the climate emergency and the challenges of regulating new technologies in a coordinated approach with other independent public and administrative authorities[116].

In addition, the ever-increasing use and spread of voice assistants in our daily lives, in both public and private spaces, raise economic and competition issues. We thus need to consider the possibility of alternatives to the already existing offerings of the tech giants, whether American or Chinese. The predominance of these players is clear in the news: the Djingo speaker, a joint Orange-Deutsche Telekom creation, for example, integrates Amazon's assistant Alexa.

From an economic point of view, the question of the remuneration of the various actors in the production chain also arises. What is the underlying business model? In the years to come, might we have a paying model to improve the ranking of its application? The CSA-Hadopi report raises the question of the distribution of profits between application and content publishers and advertising services managed by manufacturers. Are the benefits equitably distributed among these actors? Finally, the massification of the leading market players' databases is sparking reactions and concerns. What about competition and potential abuses of dominant position?

[115] - Karen Hao, *Tiny AI models could supercharge autocorrect and voice assistants on your phone*, MIT Technology Review, october 2019, https://www.technologyreview.com/2019/10/04/132755/tiny-ai-could-supercharge-autocorrect-voice-assistants-on-your-phone/

[116] - Autorité de la concurrence, AMF, ARCEP, ART, CNIL, CRE, CSA, HADOPI, *Accords de Paris et Urgence Climatique : Enjeux de Régulation*, https://www.arcep.fr/fileadmin/user_upload/publications/cooperation-AAI/publication_AAI-API_Accord_de_Paris_052020.pdf

Interview with...
EMMANUEL VINCENT

Emmanuel Vincent *is Research Director within the Multispeech team (Université de Lorraine, CNRS, Inria)*[117]*. His research interests include hands-free voice control and ambient sound analysis. He develops data-saving and privacy-friendly artificial intelligence technologies.*
He coordinates the COMPRISE project and is one of the organisers of the VoicePrivacy challenge.

Is the exposure of users' private lives inevitable with voice assistants? Are privacy-protective implementations possible? During the drafting of this White Paper and within the framework of the CNIL-Inria partnership, LINC met with Emmanuel Vincent, an Inria researcher whose work focuses on the development of new voice interfaces that meet data protection requirements from the design stage.

A large number of studies herald adoption of voice assistants on a massive scale in the years to come. What do you think will be at stake for users when all our equipment is equipped with these?

By enabling users to express complex requests, voice assistants address the need for effective interaction with everyday Internet content, objects and services. Voice technology companies will expand the languages supported and combine voice command with analysis of other aspects of the voice (age, emotions, preferences, etc.) to better characterise the user and his or her desires. Companies of all kinds will in turn integrate these technologies into a growing number of products.

This raises many issues for citizens, user companies and public authorities. For example, supporting a language has a cost that is not always commercially viable. It is essential for cultural diversity and equal opportunities to support free software and open data initiatives, so that these technologies can become accessible to all citizens, regardless of language, dialect or accent. It is also essential that the answers given by the assistants are fair and explainable: in response to a question about a product, why showcase the websites of some brands over others?

Uses must be controlled: technology such as emotion analysis can be both beneficial for more seamless interaction at a given moment and ethically reprehensible if the emotions detected are retained for commercial profiling purposes. Even when the use is acceptable, the collection of voice data raises issues of security and confidentiality.

It is therefore necessary to anticipate future functionality and uses in order to build the appropriate legal framework and enable citizens to become informed users.

The General Data Protection Regulation (GDPR) advocates a privacy by design approach. What does such a concept look like in practice in the case of voice assistants?

According to the GDPR, voice is considered to be personal data. It conveys four types of information of a personal nature: the words pronounced, the biometric characteristics of the person who pronounced them (identity, age, gender, etc.), the way they were pronounced (emotions and pathologies reflected in the voice) and the environment in which they were pronounced (ambient noise and voices). The GDPR goes further by categorising as sensitive information biometric characteristics and words betraying sexual orientation or religious views, for example.

[117] - https://team.inria.fr/multispeech

In concrete terms, voice assistants ask users for express permission to use their voice for certain predefined purposes and offer them the possibility of accessing and requesting the deletion of stored data. This is in accordance with the law, but does not allow users to finely control the uses that are made of their data, as the predefined uses are often not as specific as informed users might wish.

Following Claude Shannon's work on information theory, research in the field of automatic speech processing dates back to the 1960s. However, it seems that combining it with privacy protection techniques is still very recent. Why is this?

Voice technologies work through machine learning from voice recordings transcribed into text form. For a long time, such data was acquired from voluntary subjects and the systems only worked reliably enough for the recognition of numbers or keywords, which is not very critical for privacy.

The boom in voice assistants is due to a combination of three factors: the emergence of more powerful learning methods, the increase in computing capacity and the explosion in the amount of data. Some companies keep all the voice commands sent to their assistant for various purposes, learning in particular. This increase in the quantity and diversity of each user's data, coupled with the increased ability to extract information from it, amplifies the risks to privacy, whether for legal or illegal use (cyber-attack): profiling, access to sensitive information, identity theft, industrial espionage, etc. Profiling is a common practice that could be developed further by cross-referencing information from multiple scenarios of use. The other risks may seem exaggerated today, but are a likely threat within a few years.

To limit these risks, other companies are choosing not to retain voice commands and to use learning data acquired from voluntary subjects, at the risk that their products will be less effective.

**You're researching the subject.
Can you tell us how you came to work on these objects and the challenges you wish to take up?**

My interest stems from the observation that, in order to achieve the economic and societal benefits expected from artificial intelligence and voice assistants in particular, we need to develop effective machine learning tools capable of best harnessing massive personal data while ensuring the preservation of privacy, fairness and other values to which our citizens are attached. The trigger came from contact with the Magnet team (University of Lille, CNRS, Inria), which designs such tools and provides formal guarantees of confidentiality, and with European companies, which have expressed interest.

Since the end of 2018, in the framework of the COMPRISE project funded by the European Union's Horizon 2020 programme, we have been designing an open source voice assistant and a learning platform based on the principle of privacy protection through[118] the user towards the learning platform, we transform the voice and replace certain words so that the user is no longer identifiable. Our first tests were an uphill struggle because modern biometric tools are extremely powerful in re-identifying the user, even after transformation. Our tools do not guarantee perfect anonymity, but provide a level of protection far superior to what currently exists.

My team also coordinates the DEEP-PRIVACY project funded by the National Research Agency, which takes an alternative approach to decentralised learning. In this approach, personal data does not leave the user's device, which provides increased protection but has the disadvantage that it can no longer be manually re-transcribed. To encourage other initiatives of this kind, we have created the VoicePrivacy Challenge, the results of which will be presented very soon[119].

[118] - COMPRISE, *Cost-effective Multilingual, Privacy-driven voice-enabled Services,* (2019 – 2022), https://www.compriseh2020.eu/

[119] - https://www.voiceprivacychallenge.org/

FOCUS ON...

Voice Assistants and Privacy Protection Research

It is only in the last decade or so that research in the speech processing and privacy fields began to be linked. Studies initially focused on the use of cryptographic techniques applied to signal processing such as secure multiparty computation, unconscious transfer or homomorphic encryption[120] [121]. More recently, in parallel with the increasing uptake of voice assistants, there has been an increase in research and publications. While the security aspects of these voice-assisted devices have already been extensively discussed above (see page 35), many projects at the intersection of speech processing and privacy protection have also been funded .

As an example, the PAMELA project funded by the French National Research Agency (ANR) aims to develop machine learning methods using local personalised models, in a decentralised and cooperative way within networks where data and learning systems are distributed[122]. One of the scenarios studied is that of a voice assistant that respects privacy (the SNIPS company was a member of the consortium). As mentioned by Emmanuel Vincent in his interview (see page 43), the ANR DEEP-PRIVACY and H2020 COMPRISE projects are along these lines, also proposing to focus on federated learning and differential privacy techniques[123] [124]. Adversarial approaches are used to modify the speech signal so that it no longer carries the user's biometric characteristics without degrading the quality of the automatic speech recognition (speech to text). The French-Japanese project VoicePersonae aims to act as a hub for many topics related to voice identity such as speech synthesis, speaker recognition, detection of spoofing attacks, media forensics, anonymisation of speech, etc.

In order to federate the research community around these subjects and to measure progress in a competitive mindset, scientific challenges have also been launched. The idea is to offer different research teams the opportunity to measure themselves against each other on the achievement of a predefined task. For example, the ASVSpoof Challenge promotes the production of countermeasures to combat identity theft in the context of voice authentication[125]. The VoicePrivacy Challenge aims to develop anonymisation solutions to remove personal information from voice signals[126]. Similarly, the ANR DAPCODS/IOTics project is looking at the connected objects of the smart home from a data protection perspective. This work focuses in particular on the means for controlling the home, first and foremost the voice assistant of his/her smartphone and smart speaker. Indeed, it is generally these means of control that determine the nature of the data exchanged (which sometimes go beyond the data collected by the object), its volume and the actors involved, which also raises questions of sovereignty[127].

Finally, the HUMAAINE Chair (which stands for human-machine affective interaction & ethics) studies human-machine interactions and relationships in order to audit and measure the potential influence of affective systems[128]. The work focuses on the detection of social emotions in the human voice and, using the contributions of behavioural economics highlighted by Nobel Prize winner Richard Thaler, on the study of manipulations in spoken and audio language that are intended to induce changes in the behaviour of the human interlocutor. The end goal of this scientific work is to enable the development of ethical systems by design.

[120] - R. (Inald) L. Lagendijk, Zekeriya Erkin et Mauro Barni, *Encrypted Signal Processing for Privacy Protection*, IEEE Signal Processing Magazine, january 2013.
[121] - Manas Pathak, Privacy-Preserving Machine Learning for Speech Processing, Springer, 2013.
[122] - PAMELA, *Personalized and decentralized machine learning under constraints*, (2016 – 2020), https://project.inria.fr/pamela/
[123] - DEEP-PRIVACY, *Distributed, personalised learning, preserving privacy for speech processing*, (2018 – 2022), https://anr.fr/Projet-ANR-18-CE23-0018
[124] - COMPRISE, *Cost-effective multilingual, privacy-driven voice-enabled services*, (2019 – 2022), https://www.compriseh2020.eu/
[125] - https://www.asvspoof.org/
[126] - https://www.voiceprivacychallenge.org/
[127] - DAPCODS/IOTics, *Data protection of connected devices and smartphones*, (2017-2020), https://project.inria.fr/iotics/fr/
[128] - http://humaaine-chaireia.fr/

USE CASES: GDPR IN PRACTICE

This chapter sets out to study how the GDPR can be applied in several contexts in which voice assistants can be used, whether as a personal assistant (on a smartphone for example) or as any device that can be integrated into a more collective framework (e.g. component of the home: connected TV set, connected refrigerator, living room speaker, etc.).

The objective is to provide home and business users, as well as assistant designers and developers of applications intended for these assistants with examples to follow in order to comply with the various aspects of the regulations.

The scenarios presented below are fictitious cases inspired by what is currently offered on the voice assistants' market. They do not constitute a complete and exhaustive analysis of how the GDPR should be applied in the context of voice assistants, but do provide avenues for analysis and reflection. They are based on premises that do not necessarily reflect the modus operandi of all voice assistants. For example, these are grounded in the assumption that the main speech and action processing functions are performed online, on systems controlled by solution providers, reflecting the operation of most speech assistants currently available to the general public – but other processing models are possible.

Three use cases are presented to cover different issues:

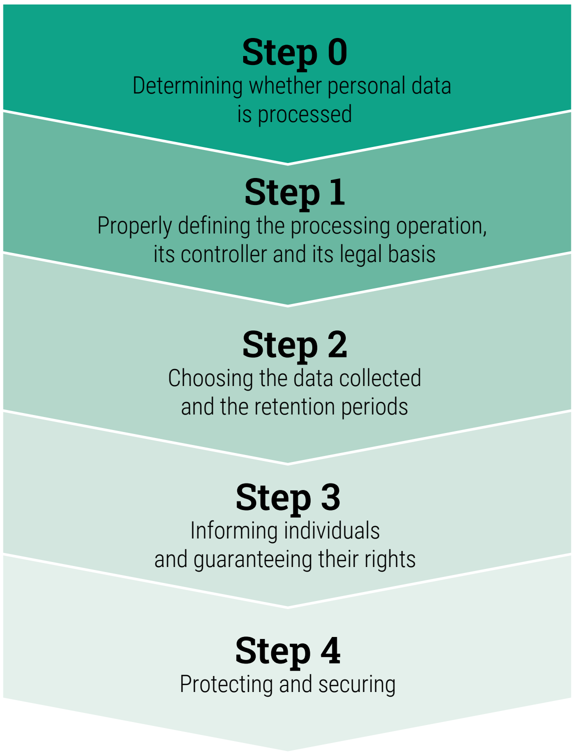
USE CASE NO. 1:
Using the basic functions of your voice assistant

USE CASE NO. 2:
Using a banking application via a voice assistant

USE CASE NO. 3:
Re-using the data collected by the voice assistant for service improvement purposes

Indeed, a distinction must be made between questioning the assistant on a general matter for which it consults publicly accessible resources on the Internet, the use of an application or the management of connected equipment through the assistant. The first use case does not require, beyond creation of a user account, the installation and/or use of a specific application. Only one player is involved: the designer of the assistant. Conversely, in the second and third use cases, several actors may be involved, namely the designer of the assistant, the developer of the application for the purchased product and/or the supplier of the connected equipment (e.g. a connected light bulb).

Thus, the compliance of the processing of the voice assistant can be analysed by following four main steps (step 0 being a prerequisite):



Step 0
Determining whether personal data is processed

Step 1
Properly defining the processing operation, its controller and its legal basis

Step 2
Choosing the data collected and the retention periods

Step 3
Informing individuals and guaranteeing their rights

Step 4
Protecting and securing

Regarding step 0, the first question to be addressed is whether the voice assistants process personal data. Indeed, one might wonder whether the collection of single sentences does indeed fall within the scope of personal data protection. The answer is dealt with directly in this introduction and is obviously: yes. As described in Chapter I *What's the story behind voice assistants?* the voice is undeniably personal. In addition, the questions asked or even the simple activity recorded by the voice assistant during its wakefulness phase can reveal personal information about the person such as location, preferences, hours of presence, etc. Finally, the transmission of the data recorded by the assistant is accompanied by personal data, either indirectly identifying (user or equipment number, IP address, pseudonym, etc.) or directly identifying (account data, email address, etc.).

The key concepts of GDPR

Collecting and processing personal data requires the key principles of data protection to be followed and their rights to be respected. In particular, individuals must be informed of the use made of their data. As a data controller, or as a data processor, it is necessary to take measures to ensure that the use of such data respects the privacy of the data subjects. The following principles apply to any processing of personal data.

Purpose & Status of the Actors: this is the main goal of the use of personal data. In fact, it is obligatory to define beforehand the precise purpose for which the data is collected or processed. The purpose is to be respected throughout the life cycle of the data. A controller is the legal (company, municipality, etc.) or natural person who determines the purposes and means of processing.

Legal Basis: there must be a legal basis for each purpose: contract, consent, legitimate interest, legal obligation, public interest mission or protecting vital interests. The details of choice and application of each of them are to be found on the CNIL website[129].

Accuracy, Proportionality & Data Minimisation: only information that is adequate, relevant and strictly necessary in the light of the objectives previously set shall be used. The data must be accurate and kept up to date. Inaccurate data must be corrected or deleted.

Retention Period Limitation: the retention period must be determined beforehand and respect a principle of proportionality and balance which depends on the aims pursued. It may not be unlimited and must last only as long as necessary to achieve the objective (purpose) previously set and brought to the attention of the data subjects, in accordance with any other legal obligations (legal framework of the banking sector, insurance, health, etc.).

Security: it is mandatory to take appropriate security measures, both IT and physical, to guarantee the integrity, availability and confidentiality of personal data. They should be adapted to the sensitivity of the data and the risks to individuals in the event of an incident.

Information & Transparency: it is essential to provide information that is concise, transparent, intelligible and easily accessible (see Data & Design box on page 69). Transparency is the basis of the contract of trust that binds organisations with the people whose data they process. It is also one of the fundamental rights of individuals (see below).

Data Control & Risk Identification: the sharing and movement of personal data must be regulated and contractualised, in order to ensure its protection at all times. Specific measures may apply in cases where large volumes of data or sensitive data are processed or where the processing of data may have particular consequences for individuals. These measures include the Data Protection Impact Assessment (or DPIA), which is mandatory in some cases (the list of criteria and the list of types of operations concerned are available on the CNIL website[130]) - see the box on the PIA methodology applied to the area of connected objects on page 56.

[129] - CNIL, *Les bases légales*, https://www.cnil.fr/fr/les-bases-legales

[130] - CNIL, *What you need to know about the data protection impact assessment*, https://www.cnil.fr/fr/ce-quil-faut-savoir-sur-lanalyse-dimpact-relative-la-protection-des-donnees-aipd

Protection of Sensitive Data: this concerns information revealing alleged racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership, as well as the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning the sex life or sexual orientation of an individual. Such data is subject to a regulation: it may not be collected or processed except in exceptional cases (in particular with the explicit consent of the individual - see page 53 and the CNIL website for more information[131]).

Rights of data subjects: data subjects have rights in order to retain control over their data. The controller must explain to them how to exercise these rights (who to contact and how, etc.). When exercising their rights, individuals must receive a response within one month. There are eight such rights:

- **Right to information:** allows the individual to be aware of the processing of personal data concerning him/her and to exercise his/her other rights.

- **Right of access:** allows the individual to know which of his/her data is being processed and to obtain communication of it in an understandable format. It also makes it possible to check the accuracy of the data and, if necessary, to have it corrected or deleted.

- **Right to rectification:** allows corrections of inaccurate data (e.g. wrong age or address) or data to be added (e.g. address without apartment number) in connection with the purpose of the processing.

- **Right to object:** allows the individual to object to his/her data being used by an organisation for a specific purpose. Such objection must be founded, except in the case of marketing, which may be objected to without legal justification.

- **Right to erasure:** allows the individual to delete his/her data (under certain conditions).

- **Right to data portability:** allows the recovery of data provided to an organisation in the context of the use of its service, in a commonly used digital format. This recovery of data can either be for personal use or for transfer to a third party.

- **Right to human intervention:** every individual has the right not to be subject to a fully automated decision where it produces legal effects or significantly affects him or her. Such a decision can only be triggered under the conditions specified in the GDPR and while preserving the right of the data subject to obtain human intervention, to express his/her views and to challenge the decision.

- **Right to restriction of processing:** this can be obtained when one of the other rights mentioned above applies (rectification, objection, etc.). If the accuracy of the data used is disputed by the body or if the individual objects to the processing of his or her data, the law allows the body to check or examine the request within a certain period of time. During this process, the individual has the opportunity to ask the body to freeze the use of his/her data.

[131] - CNIL, *Definition – Donnée sensible*, https://www.cnil.fr/definition/donnee-sensible

USE CASE NO.1:
Using the basic functions of a voice assistant

The first known use of the voice assistant is to respond simply and quickly to recurring functional needs. Thus, for the so-called "mainstream" voice assistants Chapter I.4 *What strategies for the designers of voice assistants?* a user can query his assistant on various topics as s/he would using a web search engine. For example, s/he can check the weather, find out the shortest route to work or ask to turn on the radio.

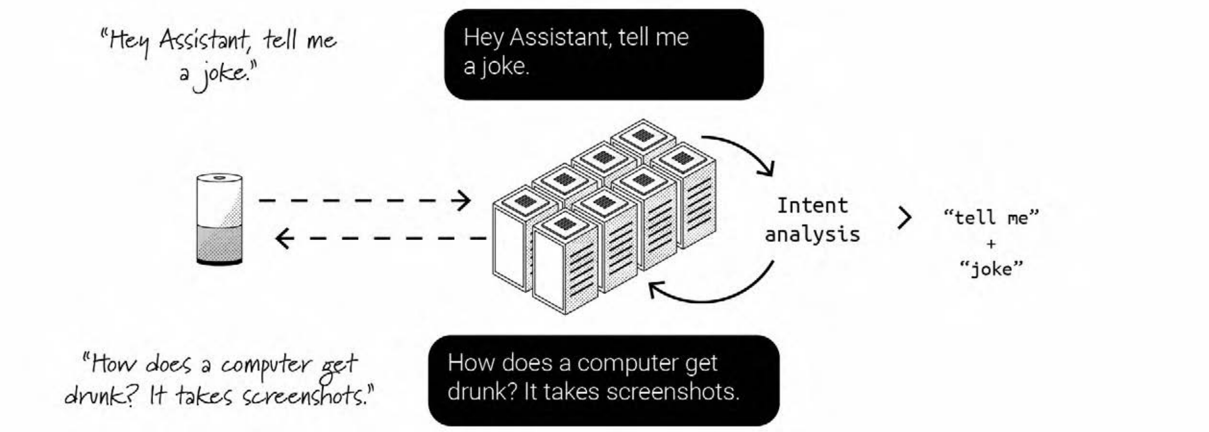

"Hey Assistant, tell me a joke."
Hey Assistant, tell me a joke.
Intent analysis
"tell me" + "joke"
"How does a computer get drunk? It takes screenshots."
How does a computer get drunk? It takes screenshots.

Figure 4
Implementation of the basic functionality of a voice assistant

In practice, as shown in Figure 4, a voice assistant operates as follow (more details in the infographics on page 14):

- Capturing the user's query following the activation of the assistant;
- Transmitting voice signal to the assistant designer's servers for automatic speech transcription, interpretation of the command and identification of a suitable response;
- Remote triggering of an answer or an action (telling a joke in this example) which is ultimately executed via the equipment integrating the assistant.

In the majority of cases, and before the voice assistant can operate, the user must have an account associated with the device. S/he will have to create a specific account or the assistant designer will allow him/her to link an already existing account to use the product directly.

NB:

A user account is not technically necessary to use the voice assistant when accessing publicly available information online (weather, news, etc.), just as it is not technically necessary to create one to be able to browse the web.

Case Study 1 Warnings

The study of this "generic" case (ask questions, ask for the weather, etc.) provides the legal basis for the relationship between the user and the designer of the voice assistant. The main principles (minimisation, restriction, legal basis, information, etc.) remain for the other two use cases presented afterwards: only the specifics relating to each use case will then be developed.

Step 1:

Properly defining the processing operation, its controller and its legal basis

Defining the purpose and status of actors

In this scenario, the processing concerns both the user account data, the commands addressed by the user to the assistant (the voice signal and its transcription into text form) and the data necessary to process the request (preferences, location, date and time, etc.).

As defined in the GDPR, the controller is the legal person (company, municipality, etc.) or natural person who determines the purposes and means of a processing operation, i.e. the objective and the way to achieve it. Here, the designer of the assistant is therefore the controller insofar as it determines the purposes (the provision of the voice assistance service) and the means (processing via the assistant linked to a user account).

Specifying the legal basis for the data processing operation

The definition of the legal basis allows the designer of the voice assistant to determine the legal grounds of the data processing operations. In this scenario where the use of the assistant is limited to the functionalities provided by the designer, the data processing described above is necessary to perform the service requested by the user via the voice assistant. Therefore, its legal basis could be the performance of a contract to which the user, as the data subject, is party (Article 6(1.b) of the GDPR).

In this scenario of use, the legal basis linked to the consent of the data subjects seems relatively unsuitable. To be valid under the GDPR, consent must be freely given, specific, informed and unambiguous. In the case of voice assistants, lack of consent would result in the inability to use their services. For example, the lack of consent to voice processing would not allow automatic transcription in this case. This negative consequence could weigh in the data subject's decision to give or not to give his/her consent, which would then not be free.

FOCUS ON...

The data necessary for the performance of the contract

The contract may only provide a valid basis for data processing if this is objectively necessary for the performance of the contract. It is not sufficient that the data processing is mentioned in contractual clauses or in general terms and conditions of use. In other words, the data processing must only enable the organisation to provide the product or service desired by the user, and must not have any other purpose allowing, for example, the pursuit of distinct or exclusive interests of the designer of the voice assistant (for more information, see the dedicated page on the CNIL site[132]). If the data is used for other purposes, the legal basis for those purposes will have to be different.

> «
>
> ***It is not always technically necessary to go through a user account in order to use a voice assistant.***
>
> »

[132] - CNIL, *Le contrat : dans quels cas fonder un traitement sur cette base légale ?*, 2020, https://www.cnil.fr/fr/le-contrat-dans-quels-cas-fonder-un-traitement-sur-cette-base-legale

FOCUS ON...

Advertising profiling and the ePrivacy Directive

As seen in Chapter I.4 *Which strategies for voice assistant designers?*, some designers of voice assistant may wish to use the history of requests that the user asks the assistant to create and populate a profile (customise the services offered by the designer, implement more targeted advertising campaigns, offer more detailed sales proposals, etc.). In application of the principle of transparency, such use of data must necessarily be brought to the attention of the data subjects and may under no circumstances take place without their knowledge. This will be a new processing operation distinct from that necessary for the operation of the service, which cannot be based on the legal basis for the performance of the contract like the latter and whose compliance with the GDPR will have to be analysed separately.

It is also necessary to apply the provisions of the European ePrivacy Directive (Directive 2002/58/EC). Indeed, Article 5(3) of this directive provides that any reading or writing operation on a user's device, through a telecommunications network open to the public, may only take place with the user's consent, unless this operation is strictly necessary for the provision of a service explicitly requested by the user or is explicitly intended to carry out the transmission of an electronic communication.

The definition of "user device" includes voice assistants. If the data entered or stored even temporarily (username, voice recordings) in the device are accessed from remote servers to inform ad profiling, consent is required.

As such, in addition to the question of the compliance of the general processing operation with the provisions of the GDPR, the question of the purposes of the particular operation of reading the data in the assistant should also be raised. Thus, if the latter is necessary for the provision of the service explicitly requested by the user, consent is not required. Conversely, if the purpose of these operations is to enhance or create an advertising profile, which by its nature is not necessary for the provision of the service, then the consent of the user must be obtained for that particular purpose.

It should be noted that devices installed in a private home, the operation of which does not involve the transmission of data to the outside world (a voice assistant that would perform all operations locally, without requiring an exchange with a remote server and therefore transmission of data to a controller) can potentially benefit from the household exemption, as provided for in Article 2(2)(c) of the GDPR. Such devices, such as a voice assistant that would only allow a user to turn on or off a household appliance, would fall outside the scope of the GDPR. By focusing on this type of architecture, voice assistant manufacturers can reduce the risk of personal data breaches, while easing their legal obligations.

Step 2:

Choosing the data collected and the retention periods

Applying the principles of accuracy, proportionality and data minimisation

In accordance with the principle of data minimisation, only information strictly necessary for the provision of the service should be processed. Using the features offered by the assistant obviously requires voice processing for wake word detection, automatic speech transcription and command analysis and interpretation. In addition, if the voice assistant is connected to an account, only data that is essential for the operation of the account and its interaction with the assistant needs to be processed. This may include identification data (surname, first name), where a pseudonym may be sufficient, and authentication data (username and password). In this case, the recordings and transcriptions of commands spoken by people in the assistant's environment are sometimes accessible from the history of the user space, which traces all the searches and questions recorded by the assistant. Technical information (e.g. IP address) or information associated with the equipment in order to verify its conformity (serial number) may also be collected, for example for system updates or product safety purposes. Some optional services may require additional information – for example, the user may choose to provide their postcode to receive relevant weather or traffic information.

As a general rule, queries made by the user to his/her assistant should not be used to infer information about the user that could fall under the category of "sensitive" data: alleged racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership, data concerning sex life or sexual orientation. In particular, any inference of the type "has searched for local Mass times = Catholic or believer", categorisation or creation of segments on the basis of such data, for the purpose of creating a profile must, above all, serve a purpose whose legitimacy will have to be demonstrated and, in any case, will require the prior collection of the explicit, specific and informed consent of the user.

Securing sensitive data: the special case of biometric data

Some voice assistants offer the user an option to identify themselves from their voice in order to access a service that is different from other users, such as other members of the household. Thus, even if several people are associated with the same device, they will each have access to the information concerning them (e-mail, calendar, customer account, etc.)

In practice, such functionality requires speaker recognition, i.e. the application of biometric processing to the user's voice. Thus, samples of his/her voice are collected to create a biometric template or template that uniquely identifies him/her so that s/he can be recognised when the assistant is called upon at a later date.

When used to identify a person, biometric data such as the voice template is qualified as sensitive data within the meaning of data protection legislation (see Chapter I.1 *What's so special with the voice?*). The GDPR strictly regulates this by prohibiting such processing in principle, except in certain specific cases. This includes the case where the user has explicitly consented to it in a completely free and informed manner. In order to give the user real freedom of choice, the designer of the assistant must offer an alternative authentication or identification method to biometrics, without any additional constraints.

Pursuant to the principle of data protection by design and by default, the CNIL, as well as the European Data Protection Board (EDPB)[133], recommends that biometric data be stored on a device under the exclusive control of the user (which in this case may be the object embodying the "voice assistant"), in the user's hand and adequately secured[134]. This mode of operation should always be favoured over a method of storing biometric information (voice template) on a remote server.

Detecting the voice of the right speaker also involves comparing it with that of other people in the assistant's vicinity. In other words, the speaker recognition functionality implemented in voice assistants may require the voice biometrics of people speaking in the household to be recorded, to allow the user's voice characteristics to be distinguished from those of the person who wishes

[133] - EDPB, *Guidelines 3/2019 on processing of personal data through video devices*, 10/07/2019, https://edpb.europa.eu/our-work-tools/public-consultations/2019/guidelines-32019-processing-personal-data-through-video_en

[134] - CNIL, *Biométrie dans les smartphones des particuliers : application du cadre de protection des données*, https://www.cnil.fr/fr/biometrie-dans-les-smartphones-des-particuliers-application-du-cadre-de-protection-des-donnees

to be recognised. Biometric identification may therefore have the consequence of subjecting uninformed persons to biometric processing, by registering their template and comparing it with that of the user wishing to be recognised. In order to avoid such collection of biometric data without the knowledge of the data subjects while allowing a user to be recognised by the assistant, solutions based on the user's data alone should be given priority. In concrete terms, this means that biometric recognition is only activated at each use at the user's initiative, and not by a permanent analysis of the voices heard by the assistant. So, for example, if the user wishes to set up biometric authentication for access to certain protected data such as his/her bank account, the voice assistant could activate speaker verification, when s/he launches the banking application only, and verify his/her identity in this way.

Limiting data retention periods

In accordance with the principle of storage limitation (Article 5(1.e) of the GDPR), data must not be stored longer than necessary for the provision of the service. Therefore, the information should be deleted by the controller as soon as the service has been rendered, for example, as soon as the answer to the question asked has been given by the assistant.

Sometimes, the service provider offers the user the possibility of keeping the history of his/her requests on his/her account. In this case, the data kept to be made available to the user must not be processed for any other purpose than that of providing the service. Otherwise, it will be a new data processing operation, the lawfulness of which will again have to be assessed beforehand in the light of the rules of the GDPR (see Step 1). If the history is kept only to be made available to the user, the user must be able to delete it easily and at any time.

When the data is no longer necessary for the provision of the service or when the user deletes his/her account, the data controller must in turn permanently delete the information it holds on the user, subject to any legal obligations requiring it to archive such information. The deletion of data reduces the potential risk of misuse. For example, data collected for the provision of the service could be used for other purposes such as marketing, improvement of the service or enrichment of the customer profile without the user being aware of this or without the applicable provisions being complied with (e.g. information to users, consent of individuals for certain purposes, etc.).

Step 3:
Informing individuals and guaranteeing their rights

Implementing the principles of information and transparency

There is a single controller who is responsible for the obligation to provide information and for managing requests to exercise users' rights. In this scenario of use, it is the responsibility of the designer of the voice assistant to ensure that the prior information is provided in clear and simple terms before the processing operation is carried out and at the latest at the time of data collection (i.e. when the assistant is started). Different media may be used by providers of these services, cumulatively: instructions for use of the assistant, pre-recording of explanatory voice messages on the assistant which are sent until the user confirms that they have been taken into account, information during the account creation procedure, provision of a privacy protection policy from an easily accessible and identifiable online space (the latter method can be used in addition to more direct and prior information via the abovementioned media), etc. In particular, the controller must explain to the user the purpose of the collection of his/her data, which is required for the proper functioning of the voice assistant, how to keep control over his/her data, including through the exercise of his/her rights, etc. If possible, the data controller may rely on the use of a companion screen, for example through a dedicated application, for the presentation of this information as well as for the collection of the user's actions (acceptance of the terms and conditions and the privacy policy, device settings, etc.).

FOCUS ON...

Voice information

Since the interface is primarily voice-based, the designer could usefully provide for the possibility of voice information through the assistant satisfying the GDPR requirements. This type of information is particularly relevant for people who use a voice assistant and are unable to use a written medium. In order to avoid a lengthy reading of the terms of use and privacy policy, their main aspects can be integrated into a short general presentation and then complemented by question/answer mechanisms allowing the user to access the information that most interests him/her.

Ensuring that data subjects' rights are respected in practice

The user has rights allowing him/her to keep control over personal information. Their existence and the manner in which they are to be exercised must be brought to the user's attention by the data controller.

The exercise of the right of access enables the user, on the one hand, to know what data are held on him/her and to obtain communication of such data in an understandable format and, on the other hand, to obtain information relating to the processing of such data: the purposes for which it is processed, the recipients of the data, its retention period, etc. In particular, the communication of data allows the user to check its accuracy and, if necessary, to have it rectified or to request its deletion (*see Key concepts of the GDPR*).

Where possible, the controller may allow the user direct access to his/her data. Many assistant designers thus offer users access to a history of their interactions with the voice assistant. It should be noted that simply referring users to such a history does not appear to enable the data controller to meet all its obligations under the right of access, as the accessible data generally represents only part of the information processed in the context of providing the service.

If the user finds that the information concerning him/her is not accurate, s/he can rectify it from his/her user account or ask the data controller to modify it; s/he can also request its deletion.

Finally, the user must be offered the possibility of exercising his/her right to data portability, in the context of the provision of the service by the voice assistant and when such data is collected on the legal basis of the contract or consent. This right allows the user to retrieve for his/her personal use, on the one hand, the data s/he has communicated in the context of the creation of his/her user account (surname, first name, etc.) and, on the other hand, the data produced by the use of the service (e.g. history of voice interactions, etc.). In practice, the data will have to be provided in a suitable and documented format, i.e. one that can be interpreted by a computer and with no restrictions on use. This can for example be done using an open format (XML, JSON, CSV, WAV, etc.), supplemented by any metadata useful for its interpretation. In order to facilitate the exercise of this right, the data controller may offer users the possibility of downloading their data directly from their user space.

Step 4:
Protecting and securing

Guaranteeing data security

Voice assistants being in permanent standby mode, they can activate and inadvertently record a conversation as soon as they assume to have detected a wake word. It is therefore possible that private, intimate or confidential conversations or even sensitive data may be captured without the user's knowledge, such as data relating to the state of health of a family member or banking data. This is all the more true when the voice assistant is embraced as a central feature of the household, accessible to all its members or to external persons (friends, cleaners, technicians, etc.).

These characteristics, which are inherent in the very operation of this type of service, require the implementation of enhanced security measures to protect the intimate space of users and third parties. These measures concern several aspects which the controller is obliged to cover: management of authorisations, incident management, server security, supervision of data maintenance and destruction, management of processors (see scenario no.2), secure exchanges (encryption of communications), etc. More details on these basic precautions that must be implemented systematically and on how to implement them can be found in the CNIL's personal data security guide[135].

[135] - CNIL, *Security of Personal Data*, 2018 edition https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle_gb_web.pdf

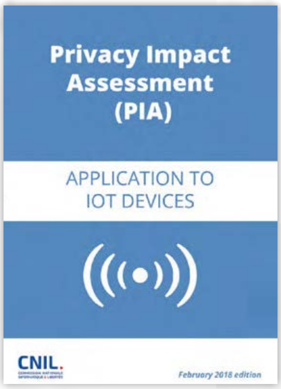
Strengthening data control and identifying risks

A Data Protection Impact Assessment (DPIA) aims to build and demonstrate the implementation of privacy protection principles. Under the GDPR, conducting a DPIA is mandatory if the processing is likely to create high risks to the rights and freedoms of data subjects. Therefore, due to the sensitivity of the data that can be processed when using a voice assistant, the DPIA may be a prerequisite.

This may indeed be the case if the proposed processing operation is included in the list of types of operations for which the CNIL has deemed it mandatory to carry out a data protection impact assessment[136] or if the treatment meets at least two of the nine criteria from the guidelines issued by the WP29 (or Article 29 Working Party)[137], the bringing together of European data protection authorities (replaced since May 2018 by the European Data Protection Board or EDPB). These criteria are: assessment/scoring (including profiling), automatic decision making with legal or similar effect, systematic surveillance, collection of sensitive data or highly personal data, large-scale collection of personal data, data cross-checking, processing of data of vulnerable persons (patients, elderly, children, etc.), innovative use (use of new technology), exclusion from the benefit of a right/contract.

FOCUS ON...

The PIA framework for connected objects



Privacy Impact Assessment (PIA)

APPLICATION TO IOT DEVICES

CNIL.

February 2018 edition

To assist companies in this approach, the CNIL has produced DPIA guides[138] taking into account the requirements of the GDPR. The methodology complies with the criteria set out in the EDPB guidelines and is also compatible with international risk management standards. In addition, in order to facilitate the implementation of DPIA, the CNIL has developed the PIA software[139], available in English and French. It is designed to facilitate and support the conduct of such an impact assessment.

More directly operational for people developing, integrating or deploying voice assistants, the CNIL has also published a version of the PIA methodology applied to connected objects[140] also known as PIAF – *Privacy Impact Assessment Framework*.

[136] - CNIL, *Liste des types d'opérations de traitement pour lesquelles une analyse d'impact relative à la protection des données est requise*, https://www.cnil.fr/sites/default/files/atoms/files/liste-traitements-aipd-requise.pdf

[137] - G29, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation (EU) 2016/679, 2017 https://www.cnil.fr/sites/default/files/atoms/files/wp248_rev.01_fr.pdf

[138] - CNIL, *Privacy Impact Assessment (PIA) guides*, https://www.cnil.fr/en/PIA-privacy-impact-assessment-en

[139] - CNIL, *The open source PIA software helps to carry out data protection impact assesment*, https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment

[140] - CNIL, *Privacy Impact Assessment (PIA) - Application to Connected Objects*, https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-piaf-connectedobjects-en.pdf

USE CASE NO. 2:
Using a banking application via a voice assistant

As discussed in Chapter I.4 *What strategies for voice assistant developers?* Some voice assistant developers allow third-party organisations to develop their own applications that can be directly queried via the assistant. Thus, more and more of these applications, accessible from an app store, are becoming available for various uses. For example, it is possible for the user to install a dedicated banking application that allows him/her to access certain services via his/her assistant.

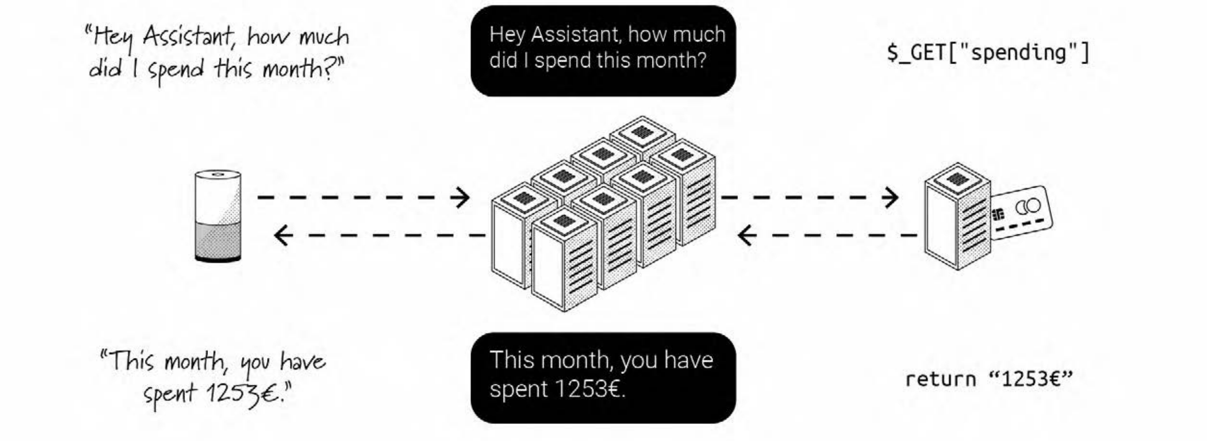

"Hey Assistant, how much did I spend this month?"
Hey Assistant, how much did I spend this month?
$_GET["spending"]
"This month, you have spent 1253€."
This month, you have spent 1253€.
return "1253€"

Figure 5

Using a third-party application via the voice assistant: consulting a bank account

In practice, as shown in Figure 5, the bank is queried by the voice assistant in the following way (more details in the infographics on page 14):

- Capture of the user's query following the activation of the assistant. The request specifies that the user wishes to access the service of his/her bank by saying the sentence "connect to (name of the bank)"; only the specifics relating to this particular scenario are presented here.
- Switches to the environment set up by the bank and sends the request to the bank's information system via an application programming interface (API) system;
- Integration of the information communicated by the bank in the response formulated to the user by the designer of the assistant.

To have access to this service, the user must in most cases have a user account associated with his/her assistant. Once this has been created (see scenario no.1), s/he will proceed to pair his/her bank account with the voice assistant. To do so, s/he must, for example:

1. Sign in to his/her bank's customer account;
2. Authorise pairing between the two accounts and accept the general terms and conditions of use of the service;
3. Once the pairing is done and an access token is issued, the user client of the bank can use the service as s/he wishes, (by saying the activation sentence of this service: "Login to (bank name)".

Case Study 2 Warnings

The study of this case provides the legal basis for the relationship between the user, the designer of the voice assistant and the third-party application developer. The question of the distribution of responsibilities is studied in particular, through the example of a banking application. The analysis developed in scenario no. 1 on the creation of the user account for the assistant configuration obviously applies for this first step. More generally, the principles set out in the introduction to this chapter and detailed in scenario no. 1 remain applicable: only the specifics relating to this particular scenario are presented here.

Step 1:

Properly defining the processing operation, its controller and its legal basis

Defining the purpose and status of actors

In this scenario, two actors are involved in the processing of personal data: the designer of the assistant and the developer of the banking application.

Determining the role of the actors is a prerequisite. It makes it possible to identify the obligations incumbent on each of the parties involved in the processing of personal data, the provision of information to individuals and the management of requests for the exercise of rights, to notify security breaches leading to a data breach, etc.

In the scenario presented, the bank is the data controller for the provision of the service since it determines the purposes and essential means of processing related to the application allowing interaction with the assistant. Indeed, it offers a dedicated application that allows the user, a customer of the bank, to manage his/her accounts remotely. In addition, it decides on the means of processing even though the processor, the designer of the assistant, plays an important role in determining these means. For example, it can operate the development platform that allows third-party applications to be integrated into the assistant. It therefore sets the framework and conditions to be respected by application developers and publishers. Pursuant to the provisions of Article 28 of the GDPR, the processor must offer his/her client sufficient guarantees that appropriate technical and organisational measures are implemented so that the processing meets the requirements of this regulation and ensures the protection of the rights of the data subject.

The relationship between the data processor (the designer of the voice assistant) and the application developer is therefore not insignificant. Indeed, the processor must indicate precisely in the contract the guarantees it implements and not refer to general principles, in particular, with regard to the security measures applied, the retention periods or data subjects' rights. In practice, guarantees must be provided as to the processing of the data passing through the assistant and necessary for the proper functioning of the banking application. In particular, the processor must undertake to process the data only for the bank's needs, under satisfactory security conditions, and to notify the bank as soon as possible of any data breach. For more information on the measures to be put in place, the parties may refer to the data processor's guide prepared by the CNIL and available on its website[141].

FOCUS ON...

The data controller and the data processor

The data controller is the entity that determines the purposes and means of the processing. The data processor shall process the data on behalf, at the instruction and under the authority of the controller. The processor may propose a solution to the controller or even influence its implementation. However, it does not decide to make use of it and does not process the data on its own behalf in the strict framework of the provision of the service.

When the user queries his/her assistant, his/her voice travels through the voice assistant designer's servers to be transcribed into text and interpreted. Then, the answer formulated by the bank is recorded in the assistant designer's information system to be synthesised. From then on, the latter can access the information circulating through its servers in order to answer the query formulated by the user. Under no circumstances may this information be used by the designer of the assistant for its own account and for its own purposes, insofar as it is acting on behalf of the bank as a processor.

Specifying the legal basis for the data processing operation

The banking application is intended for users of the voice assistant who are already customers of the institution and who have concluded an account agreement with it (for example, opening a bank account). There is therefore, independently of the purchase of the assistant, a contract between the customer and his/her bank.

[141] - CNIL, General Data Protection Regulation: a guide to assist processors, https://www.cnil.fr/fr/node/23975

As previously mentioned, in order to use the voice assistant to interact with the bank, the user will first need to associate his/her bank account with the device. The use of the service offered by the bank therefore corresponds to an additional functionality offered to the customer to consult his/her accounts and to have access to the bank's services. The processing of personal data necessary for the provision of this functionality may be based on the legal basis of the performance of the contract to which the user is party (Article 6(1.b) of the GDPR).

On the other hand, processing operations carried out on such data for a purpose other than the provision of the service requested and expected by the user must be subject to a separate analysis and have their own legal basis.

Step 2:
Choosing the data collected and the retention periods

Applying the principles of accuracy, proportionality and data minimisation

When the user chooses to rely on the services offered by the bank's application, the data collected corresponds, on the one hand, to the information entered by the user in his/her bank customer account at the time of the bank account pairing and, on the other hand, to the audio recordings, to the corresponding speech transcriptions made by the assistant designer's servers and to the intentions – the tasks that the user requests to be performed – identified in the text analysed by the assistant designer's solution (see the infographics on page 14).

Once the audio recordings have been transcribed, the text analysis is done, as in scenario no. 1, on the assistant designer's servers. It is on this technical infrastructure, in a space reserved for it, that the bank has configured the software component of its application based on basic functions made available by the designer. The tasks that the user can perform from the bank application are defined: in particular, the commands to implement them and the responses to be given once the corresponding intentions have been identified are specified. These may take the form of an oral response, an operation to be performed, or both. Please note that when the user chooses to use the services offered by the bank application, the data required for the installation of the service corresponds to the data required for the creation of the user account (see scenario no. 1).

In the event the data processor wishes to re-use the data for its own account and for a different purpose from provision of the service (for example, for the purposes of personalising advertising or improving the services of the assistant by listening to conversations), it will then have the status of controller and will have to justify a legal basis for this new processing operation, which will have to be implemented in accordance with the GDPR.

Limiting data retention periods

In accordance with the principle of data storage limitation (Article 5(1.e) of the GDPR), data must be stored for no longer than is necessary for the purpose for which it was originally collected or processed.

Consequently, the bank will have to delete the data (such as, for example, text transcripts, identified intentions as well as logs or timestamps, etc.) once it has responded to the user's request, unless it demonstrates the need to retain it in order to provide the service or to meet a legal obligation or for the purposes of proof and in accordance with the limitation periods for legal actions.

Similarly, the processor (the designer of the assistant) will have to delete the answers provided by the bank once they have been communicated to the user. This should be done regardless of whether or not the user deletes his/her activity history from the user account settings (see scenario no. 1).

Step 3:
Informing people and guaranteeing their rights

Implementing the principles of information and transparency

The bank, as data controller, must inform the user of the purpose, the data collected, the recipients, etc., prior to the implementation of the processing operation and at the latest at the time of collection of the data. This information can, for example, be delivered when the user logs in to his/her bank customer account when pairing his/her account with the voice assistant. In addition, for the sake of clarity and education, it could be useful for the bank to explain the various stages of processing from the collection phase by the processor to the bank's response via the assistant, specifying which data is accessed by each of the players, why and for how long.

Ensuring that data subjects' rights are respected in practice

In accordance with the GDPR, the user may assert all his/her rights with the data controller (access, rectification, data portability, etc.) – see scenario no. 1 and *Key concepts of the GDPR* on page 48. While it is the responsibility of the controller, i.e. the developer of the application, to provide information to the data subjects on the processing operations at the time of data collection, the designer of the assistant should, to the extent possible, assist the controller in fulfilling its obligation to respond to requests to exercise the rights of the data subjects.

Step 4:
Protecting and securing

Guaranteeing data security

Two main risks related to data confidentiality can be identified. The first relates to access to banking data by users other than the account holder, while the second relates to access by the designer of the assistant to banking data.

First of all, until the assistant has been paired to the bank's customer account, no information is transmitted to the assistant designer's servers. Next, it should be noted that the pairing that allows the user to link his/her bank customer account to his/her voice assistant, by entering a security code, is done in a single step that is permanent. The user may choose to end the use of the service by revoking the authorisation granted by deleting the link between his/her assistant and customer account from the settings of the banking application. Thus, once the pairing has been completed, any person with access to the assistant can query the assistant about the beneficiary's banking transactions without having to re-enter the security code. For example, someone close to the bank's customer user might ask the assistant a question about the details of the last three banking transactions, even though s/he is not the account holder.

It is therefore advisable to introduce a limited duration of pairing between the accounts and the assistant, regular notifications and reminders of the means of revocation of access tokens in order to protect the exchanges between the assistant designer's servers and the bank over time. In addition, a two-factor authentication, prior to connection to the banking application, can enable the user to protect access to his/her banking data. Without such authentication, and given the sensitivity and confidentiality of banking data, users are advised to switch off their assistant when other people are in the vicinity of it. This precautionary measure, which seeks to avoid the inadvertent recording of conversations, will also limit the assistant's questioning on confidential subjects (for example, "what is the bank account balance?").

The relationship with the data processor must be strictly regulated. It is necessary to document the means of ensuring the effectiveness of the data protection guarantees offered by the processor: data encryption, transmission encryption, traceability, etc. The contract with the data processor must in particular define the subject, duration and purpose of the processing and the obligations of the parties relating to confidentiality, conditions for the destruction of data at the end of the contract, notification of incidents, etc.

FOCUS ON...

Confidentiality of banking data

It is possible to distinguish between several types of banking transactions that require different levels of security, ranging from simple authentication (general account information) to much stronger authentication (request for a bank statement, credit card cancellation, transfer, etc.), depending on the level of risk presented by the transaction. To ensure the confidentiality of banking data when the user queries his/her bank, the controller should distinguish between routine management operations that require a simple verification of the identity of the customer and more sensitive operations that will entail appropriate security measures. For example, a transfer must be confirmed by another factor (e.g. sending a secret by text message or email) to authenticate the customer.

USE CASE NO.3:
Re-using the data collected by the voice assistant for service improvement purposes

For the designer, improvement of the service entails perfecting the operation of his/her voice assistant. This can be done by having a clearer idea of the uses of the device by implementing usage and operation statistics, but also by correcting the wake word detection, automatic speech transcription and natural language understanding capacities.

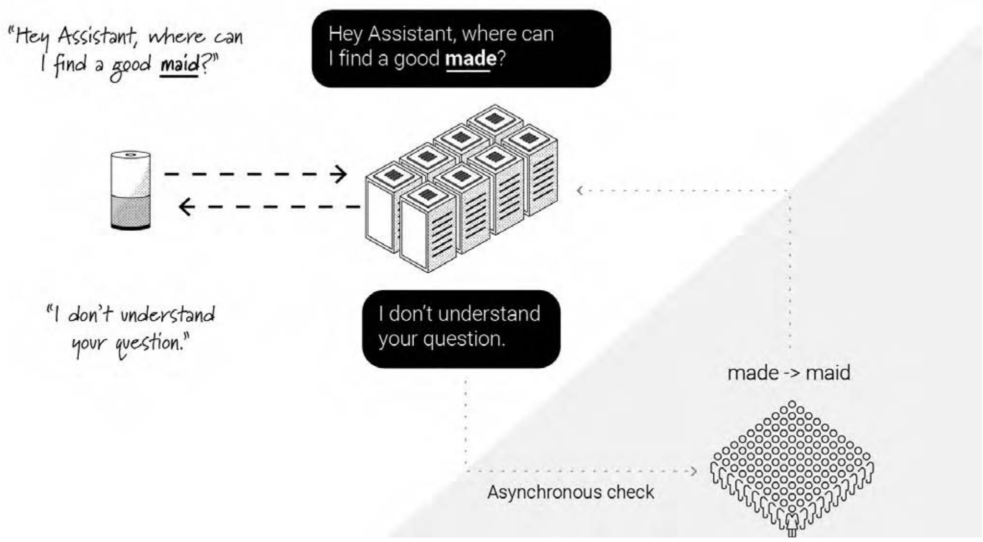

"Hey Assistant, where can I find a good maid?"
Hey Assistant, where can I find a good made?
"I don't understand your question."
I don't understand your question.
made -> maid
Asynchronous check

Figure 6

Process for improving the automatic comprehension capabilities of the voice assistant

In addition to the collection of usage data, data related to the way the user has interacted with his/her voice assistant is also collected. As discussed in Chapter II.2 *What issues for voice assistants?*, the artificial intelligence systems on which voice assistants are based require the use of learning examples. And yet, categorisation of the latter requires human supervision in order to improve the performance of the systems. Specifically, service improvement involves checking the operational functioning of the system and reporting faults so that systems are able to correct them. These operations are carried out by different people, including language analysts, who ensure the correct textual transcription according to the devices used, and data analysts, who check the meaning of past queries and their interpretation by the assistant.

In practice, as illustrated in Figure 6, the data collected by the voice assistant for service improvement purposes is reused as follows:

- Capture of the user's query following the activation of the assistant;
- Possibly, notification of a failure to provide an appropriate response to the user;
- Verification by personnel of the correct activation of the assistant, the correct transcription of the words spoken, the correct execution of the command associated with the order, etc.;
- Annotation of new learning examples to improve the performance of artificial intelligence systems.

Case Study 3 Warnings

This case study examines the lawfulness of collecting data from interactions with a voice assistant for the purpose of service improvement. More specifically, it allows to question the balance to be struck between the interests of the designer of the voice assistant and measures to protect the privacy of its users. The modalities of human listening are studied here in particular. The analysis developed in scenario no. 1 on the creation of the user account for the assistant settings obviously applies in this instance too. More generally, the principles set out in the introduction to this chapter and detailed in scenario no. 1 remain applicable: only the specifics relating to this particular scenario are presented here.

Step 1:

Properly defining the processing operation, its controller and its legal basis

Defining the purpose and status of actors

The objective is to improve the voice comprehension skills of the assistant to enable it to respond accurately to requests. Therefore, and 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 assistant, on whose behalf and for whose benefit the processing is performed.

Specifying the legal basis for the data processing operation

For the purpose of improving the service, the legal basis of the contract, referred to in scenarios 1 and 2, is not appropriate. The processing of personal data does not appear to be necessary for the performance of the service expected by the user. Indeed, the service requested by the user can be provided by the assistant without the data being used to improve the service. In this case, only the legal bases of legitimate interest or consent seem to be available. A legitimate interest of the controller may be demonstrated unless the interests or fundamental rights and freedoms of the data subject prevail. The processing operation must therefore meet the reasonable expectations of individuals and include guarantees enabling them to retain control over their data. In view of the particularly intrusive nature of listening to and analysing extracts of conversations or users' queries – extracts which may, moreover, contain sensitive data – the legal basis of free, specific and informed consent should be favoured, and over and above the enhanced guarantees to be implemented by the controller, the consent of the data subjects is the appropriate legal basis for preserving the control of individuals over their data (Article 6(1.a) of the GDPR). Access to the service provided by the assistant must not, in any event, be conditional on the user's acceptance of the use of his/her data for the purpose of improving the service. In addition, access to information collected by the voice assistant, for purposes other than the provision of the service requested by the person or the implementation of communications by electronic means, requires the prior consent of users, in accordance with Article 5(3) of the ePrivacy Directive (Directive 2002/58/EC) (see box on page 52). For more information on the conditions for collecting such consent, please refer to the CNIL guidelines on cookies and other trackers[142].

Step 2:

Choosing the data collected and the retention periods

Applying the principles of accuracy, proportionality and data minimisation

In application of the data minimisation principle (Article 5(1.c) of the GDPR), only information strictly necessary for the improvement of the service should be collected. In order to enable these improvements in the functionality of the voice assistant, including a better understanding of user requests, the data collected corresponds, on the one hand, to the users' voice recordings when the assistant is switched on and, on the other hand, to the transcribed text.

Limiting data retention periods

In accordance with the principle of limiting retention, voice recordings must be deleted as soon as the corrections necessary for the proper functioning of the device have been made. It does not seem appropriate to keep the data beyond the correction and improvement phase of the assistant, especially since it will be possible for the assistant designer to acquire new data for subsequent improvements. In order for the information to be permanently deleted, the designer of the assistant must delete it from its information system and, where appropriate, ensure that its processor has done the same (Article 28(3.g) of the GDPR).

[142] - CNIL, *Cookies and other tracking devices: the CNIL publishes new guidelines*, july 2019, https://www.cnil.fr/en/cookies-and-other-tracking-devices-cnil-publishes-new-guidelines

Step 3: Informing people and guaranteeing their rights

Implementing the principles of information and transparency

The controller must inform the user at the time of purchase of the voice assistant or its installation about the purpose of improving the service and what this means. If employees or processors are in charge of listening to the voice recordings, it will have to specify to users the nature of the listening, the length of time the recordings will be kept, the information accessed, the legal basis, etc. The data controller may carry out this information in the manner indicated in scenario no. 1 (page 50), for example by including it in the explanatory note provided for the installation of the voice assistant, or in the user account settings. The latter may or may not allow the activation of the listening carried out by employees/contractors for the purpose of improving the service.

Ensuring that data subjects' rights are respected in practice

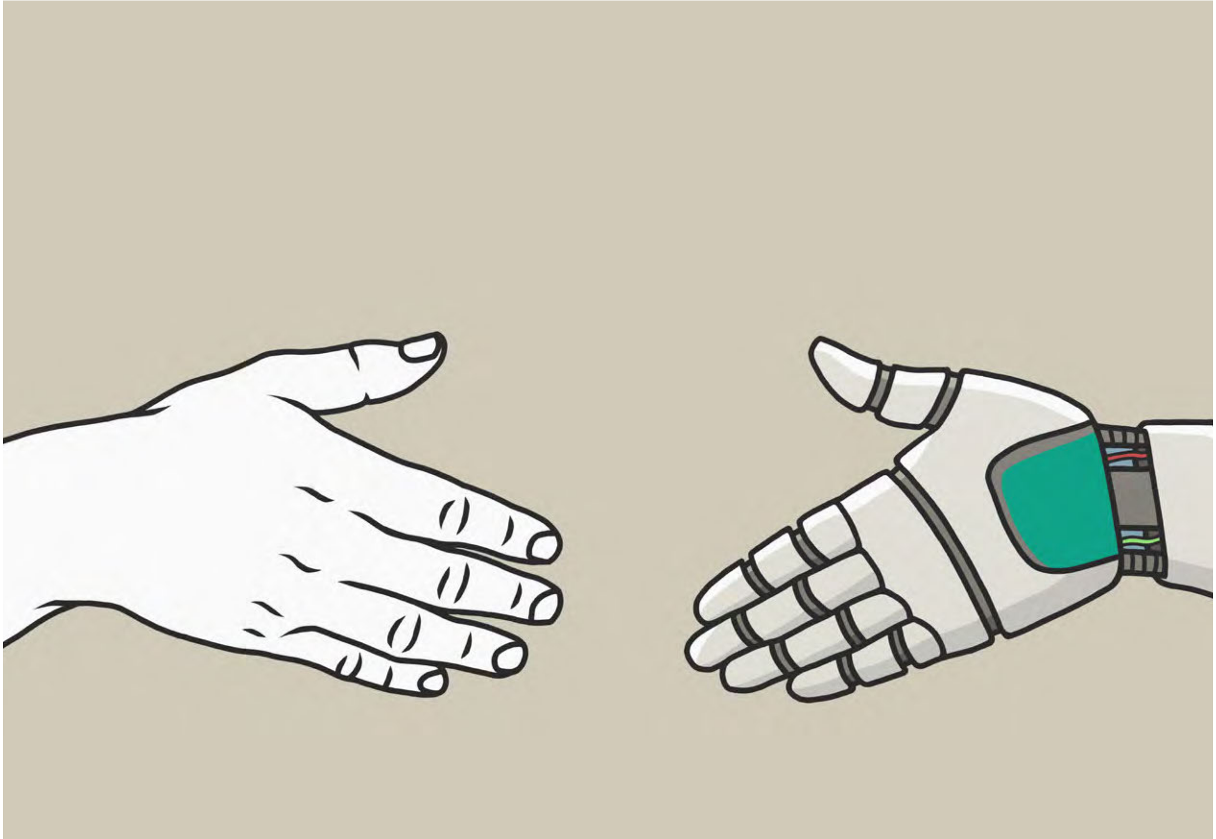
For data processing operations based on user consent, the designer of the assistant must allow the user to withdraw consent at any time. For this, a simple technical modality equivalent to that used to obtain consent must be implemented.

The data subject must also be able to exercise in a simple way and, if s/he so wishes by electronic means, his/her rights of access, erasure (e.g. if s/he has withdrawn his/her consent), restriction of processing, and right to data portability (under the conditions laid down in the GDPR) on the data used for the purposes of improving the service (see use case no. 1 - *Ensuring that data subjects' rights are respected in practice*).

Step 4: Protecting and securing

Guaranteeing data security

In the first place, as in scenario no. 1, the standard data protection measures must be implemented, as well as the measures to be taken in a contractual relationship with a data processor as identified in scenario no. 2. When the designer of the voice assistant decides to use internal employees or those of a subcontracting company to listen to the voice recordings, it must implement a number of safeguards. In particular, it must restrict access to data from voice recordings to those employees who are authorised to listen to the conversations for the purpose of analysing the replies and correcting them in order to improve the service. Moreover, only the necessary information should be accessible to these persons. For example, access to user account information is not required for the purpose of enhancing the automatic speech transcription engine. The selection of voice commands as well as the choice of assistants from which such selection is possible (through user consent, for example) must be random. Finally, the devices, from which conversation samples are listened to, must not be systematically the same



VOICE ASSISTANTS, DOING THINGS RIGHT

Voice assistants present significant privacy issues, and it is essential to remain vigilant. Thus, there are points to be kept in mind by everyone involved in the chain of responsibility of a voice assistant, from its designers to its users, including integrators, application developers and those who would choose to deploy such devices in places open to the public or places of transit (such as waiting lobbies, rental cars, meeting rooms, etc.).

Methodical and repeated questioning is indispensable in a privacy by design approach, i.e. which is aimed at implementing the appropriate technical and organisational measures to guarantee the protection of privacy and fundamental freedoms from the very beginning of the project.

In order to build the confidence necessary for users to accept devices equipped with voice assistants, the CNIL has identified four key principles:

1 - Maintaining positive friction: rather than focusing on implementing an absolutely seamless user experience, take advantage of moments of interaction (i.e. moments of choice, of settings, requiring the user's attention) to present the reality of data processing to users in an adapted manner (see box on page 69).

2 - Prefering the local to the remote: as far as possible, implement data processing modalities and capacities directly in the devices, which gives the user a good level of control over them and is a factor of confidence and acceptability.

3 - Ensuring the means of control: enable the user to understand and control the uses made of his/her data and to configure the device's operation according to his/her choices.

4 - Adapting to the voice medium: relying on audio-only interfaces raises significant challenges in terms of presenting information to the user, obtaining consent or implementing means of control. It is therefore necessary to reflect on the means to be deployed.

As presented in Chapter III *Use cases: GDPR in practice* (page 46), the use of a voice assistant must meet data protection requirements. Specifically, it is necessary to ensure that all the key principles outlined in the GDPR are met (see *The key concepts of GDPR*, page 48):

Purpose and Status of the Actors

Legal Basis

Accuracy, Proportionality & Data Minimisation

Retention Period Limitation

Security

Information & Transparency

Data Control & Risk Identification

Protection of Sensitive Data

Rights of Data Subjects

Here are some best practices for the various audiences involved in the value chain regarding the development, deployment or use of voice assistants. These must be seen as avenues for development and improvement to better protect users and their personal data. In addition, as indicated in Chapter I.2 *Voice assistant, who are you?*, depending on business models and technological choices, some players may take on several combinations of roles and thus be affected by several of the points of attention highlighted here. Compliance with this advice does not prejudge any decisions that the CNIL may take with regard to the players in this chain, particularly in the context of investigations or litigation proceedings.

FOR VOICE ASSISTANT DESIGNERS

Focusing on the software aspects, voice assistant developers are responsible for the technical implementations that will govern the operation of the voice assistants. Activation modalities, choice of architecture, data access, voice recordings management, hardware specifications, etc., it is through these design choices that the assistant's possibilities are materialised. In order to ensure that users have control and oversight over their data, seven points of vigilance must be kept in mind.

Establishing transparency as a basis for trust

As for any processing of personal data, the GDPR imposes a duty to inform data subjects of processing operations carried out by voice assistants (see *The key concepts of GDPR*, page 48). Even more than for online services, the use of a voice assistant requires users to trust a device that sometimes reveals little about how it works and whose control means are not as integrated as those used on a computer or smartphone.

Best practices

- Be transparent about the operation of the voice assistant and, in particular, about the different stages of processing from the collection phase through to the transcription of the voice into text for analysis and response to the user.
- Provide this information before purchasing the device, either by putting the information on the packaging or by making an information note available to future customers.
- Give a reminder of this information the first time the device is used, which could also be offered in audio version.
- If a lot of information is to be given, provide it in layers to prioritise the elements to be presented.
- Design the interfaces so that users can easily navigate through the different layers of information and find the ones they need at any time (during set up or later).
- Provide an oral and understandable presentation of the terms of service and privacy rules, accessible by questioning the assistant.
- Allow the user to ask questions about the personal data processing operations of the voice assistant and provide clear answers orally.
- If the designer further uses the data to improve its own services – for example by employing people to listen to and annotate recorded interactions – specifically and clearly inform the user of this use, and indicate in the management interface of the assistant the commands and recordings that have been subject to such use.

- Set the default setting of the device in its most privacy-protective operation for its user.
- If the voice assistant requires major software upgrades and/or updates, plan to contact the user to inform him/her of this and specify the nature of the changes and their consequences.
- If the assistant designer also provides a software development kit (SDK), include features and software tools that enable third-party application developers to implement the transparency requirement.

FOCUS ON...

Data and design for privacy-friendly interfaces

With the publication of its IP Report n°6: *Shaping Choices in the Digital World*, the CNIL took steps to promote the emergence of a more responsible interface design that respects data protection principles[143]. Like legal and technical issues, interface design must now be at the centre of the regulator's concerns, just as it is already at the heart of the relationship between individuals and service providers.

Following this publication, the Data & Design platform was launched[144]. This aims to create opportunities for collaboration and spaces for exchange between designers to co-construct paths that respect privacy. The goal is to incorporate these reflections in practice into the daily work of designers in order to help them explain and justify their choices and work more closely with other roles (product owners, project managers, legal departments, etc.) on the protection of personal data.

Various contents explaining and illustrating the points of the regulations on which designers can act are made available. In fact, the Data & Design platform is structured around complementary approaches relating to the explanation of key concepts of the GDPR (information to data subjects, consent and exercise of rights), the provision of case studies and the creation of spaces for exchanges on these issues both online and in physical meetings. Although the Data & Design work is not specifically aimed at voice interfaces, the elements contained therein can provide food for thought on good practices to be implemented with voice assistants.

Giving control to users

Another major requirement of the GDPR is the implementation of means to enable individuals to control the uses that are made of their data and to exercise their rights in a simple and effective way. These control and exercise modalities must be adapted to the assistant's voice interface.

[143] - LINC, *IP6 Report: Shaping Choices in the Digital World*, january 2019, https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf
[144] - https://design.cnil.fr/en/

Best practices

On the software architecture

- Promote software architectures that respect user privacy by design. For example:
  - For services that do not require remote access (alarm clock, light control, etc.) implement processing that is only carried out locally.
  - To minimise the exposure of personal data, implement as much as possible the principles of edge computing so that only strictly necessary data is transferred to centralized servers.

On the configuration methods

- Allow users to easily manage their data (listen, delete, detect abnormal usage and, if necessary, recover),
  - through a dashboard accessible via a companion screen;
  - directly by questioning the assistant by voice.
- Allow the user to fine-tune the functionalities accessible via his/her voice assistant. For example, provide a feature for automatic deletion of information once the user has received a response to his/her request, or after a period of time that s/he may determine.
- Provide seamless and easy-to-access interfaces for the management of third-party applications, either voice-based and/or relying on a companion screen, and allow the user to disable pre-installed services and applications.
- Consider, depending on the criticality of the possible applications, the implementation of filtering methods for young children that can be activated by their parents.

On the configuration modalities

- Provide the user with a means of physically disabling the device's microphone.
- Provide the user with a manual activation function, which can either trigger listening to the instructions or activate a defined period of waiting for the activation keyword.
- Indicate to the user by a sound signal the beginning and the end of the recording periods.
- Offer the user a specific voice command to deactivate the device (e.g. when there are guests, etc.).
- Consider, from the design stage, the possibility of use by dependent or disabled people. For example, a light signal indicating that the device is in active listening mode is not suitable for visually impaired people.

On account management

- Allow the association of one or more personal accounts with the assistant according to the possible uses made of it.
- Allow an account not to be associated, or a generic account to be associated when the assistant is intended for a collective or public place, or for professional use.
- Offer a private browsing mode for actions that do not require authentication, allowing a user to interact without an account being associated, nor any record of these interactions being kept.
- If several personal accounts are associated with the same device, implement reliable means of authentication for switching from one to the other and thus prevent possible identity theft.

Ensuring that the data collection is properly dimensioned

Through their interactions with them, users of voice assistants are likely to transmit a lot of information even if they do not plan to. Moreover, depending on how the devices are activated (e.g. after a wake word is spoken), inadvertent recordings may also occur.

Best practices

- Determine separate retention periods according to the type of data collected. For example, data associated with the user account can be kept longer than one-off queries to the voice assistant.
- Do not request creation of a user account if the assistant does not require it, for example if its function is to provide generic information or to program simple actions.
- Do not retain records caused by a false activation or, at a minimum, specifically identify them so that the user can be notified.

Processing biometric data

As some voice assistants are intended to be deployed in shared environments, some manufacturers offer to associate accounts for each user with it (for example, different members of a household). One option for switching from one account to another is based on speaker verification or identification. However, these are based on the use of biometric data – voice templates or models – which are considered sensitive data under the GDPR. As a reminder, the Regulation prohibits the processing of such data, with certain limited exceptions (Article 9(2)) (see Key concepts of the GDPR, page 48).

It is therefore essential 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 analysed in this way. Moreover, the user must have an alternative authentication or identification method that does not present additional constraints in order to enjoy genuine freedom of choice (see Chapter III *Use cases: GDPR in practice*, page 46).

Best practices

- For non-personal voice assistants, i.e. those that can be used by more than one person or arranged in a shared space, provide a specific keyword or question to the persons present and thereby obtain their consent to trigger biometric processing. For example, the user can say "authentication" or the assistant can ask "do you wish to be identified" and wait for a positive response to activate biometric processing.
- Keep the user's biometric template under his/her exclusive control and favour storage on a personal medium, which may be the device carrying the assistant.
- Carry out authentication/identification operations locally, i.e. directly in the device carrying the assistant.

Satisfying the security requirement

The GDPR specifies that the protection of personal data requires appropriate technical and organisational measures to ensure a level of security appropriate to the risks. The analysis of these risks is therefore a crucial step that must be carried out before designing the voice assistant. In particular, it may take the form of a formalised process known as a Data Protection Impact Assessment (DPIA). This approach, which is mandatory in some cases, and strongly recommended in others, has been specified by the CNIL in numerous tools and methods[145] (see also box on page 56).

The implementation of any processing of personal data therefore implies an obligation of security. In addition to the generic measures that can be found in the guide "Security of Personal Data"[146] and in the "Developer's Guide" (presented in the box on page 75), good practices specific to voice assistants can be specified[147].

Organising the application ecosystem

As previously presented, some assistants – especially those most popular with the general public – are positioning themselves as a platform to host third-party applications. These modes of operation need to be accompanied by specific measures and increased attention to data sharing.

[145] - CNIL, *Privacy impact assessment (PIA)*, https://www.cnil.fr/en/privacy-impact-assessment-pia
[146] - CNIL, *Security of Personal Data*, 2018 edition https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle_gb_web.pdf
[147] - CNIL, *GDPR developer's guide*, https://www.cnil.fr/en/gdpr-developers-guide

Best practices

- Choose an assistant activation method that is proportionate to the risk level of the services managed by the voice assistant (for example, a button in some cases).
- Implement a reasoned configuration for wake word detection: favour a low rate of false acceptances to avoid unexpected triggers.
- Allow the user to choose his/her wake word, with advice about the choice: the wake word must meet certain criteria (not be used too frequently in discussions, not be too similar to other words, etc.).
- Identify high-risk applications and propose security measures such as two-factor authentication (e.g. via a verification code sent by email or text message).
- Carry out a Data Protection Impact Assessment and regularly update it to ensure that the technical and organisational measures taken are in line with the risks that the data processing poses to individuals (for more details on the DPIA, see box on page 56).
- Provide mechanisms for informing and alerting the user in the event of malfunctioning of the assistant or unusual activities, including data breaches[148].

Regulating the use of data for technology improvement

Voice assistants, like other connected objects, can send personal data back to their designers for the purpose of improving services. This may include technical data used for statistical purposes (e.g. information on the use of a connected light bulb and its lifespan). In the specific case of voice assistants, the question of product improvement may also involve processing of data from voice commands, such as audio recordings or their text transcriptions.

Best practices

- Where a third-party actor uses the technological resources made available for the development of its application, contractually define the applicable rules on confidentiality and privacy in a sufficiently clear and precise manner.
- Specify the chain of responsibility involving the assistant designer and the application developer.
- Accompany application developers in the implementation of a secure service, such as generic authentication APIs and presentation of information adapted to the device.
- Limit the number of applications available by default to what is strictly necessary and encourage the installation of applications at the user's initiative, for example via an application store rather than directly from the assistant, without prior selection or specific information.
- When the assistant directly accesses a third-party application, do not share any personal data with the third party without clear information provided to the person.
- Implement a validation policy for applications deposited in the store and regularly check the store, in particular by monitoring the presence of applications with names very similar to legitimate applications.
- Provide the user with granular control tools for all installed applications and the data accessed by them, including sensitive or privacy-revealing data (health or biometric data, geolocation, search history, etc.). It should also be possible for these controls to be temporary (granting access once or for a limited period of time).
- If authorisation protocols are implemented to allow a third-party application to access a service, ensure that the access tokens used to authenticate users have a limited and reasonable lifespan and are easily revocable[148].

[148] - CNIL, *Notifier une violation de données personnelles*, https://www.cnil.fr/fr/notifier-une-violation-de-donnees-personnelles

Indeed, voice assistants rely on artificial intelligence algorithms whose performance is directly correlated to data sets used for learning statistical models. Therefore, it may be legitimate to wish to access data relating to the use of the device in real conditions in order to work on its improvement. In practice, this may mean, for example, that the designer of the voice assistant uses in-house employees or those of a subcontracting company to listen to and annotate the voice recordings so that they can be used to improve the models. Close attention must be paid to the way in which such data uses are implemented.

The rights of data subjects must therefore be respected by not processing any data for this purpose without ensuring that they are properly informed and that the legal basis for the processing is sound. Moreover, data subjects must be able to know clearly at all times whether their data is being used for this purpose and to object easily.

Best practices

- Implement strong security measures: restrict access to voice recording data to only those employees who are authorised to listen to the conversations, strong authentication, enhanced traceability measures, blocking the retrieval of audio recordings, etc.
- Do not provide any information about the user of the assistant with the recording and, in particular, do not match the recordings and transcribed files with other data that may be collected (device credentials, location, associated accounts, etc.).
- Implement measures to alter the information in audio recordings through, for example:
  - A change in the characteristics of the speaker (timbre, pitch, prosody, etc.) irreversibly;
  - A deletion/obfuscation of the information contained in the records and transcripts (surnames and first names, address, etc.).
- Sample the message concerned in several parts to be analysed by different people.
- Limit the length of listening time of user interactions to a few seconds per sample.
- Ensure that the devices from which conversation samples are listened to are not always the same.
- In the event employees of a subcontracting company are called on, provide in the subcontracting contract for all the necessary security guarantees (requirement to include a confidentiality clause in the employment contracts of the staff concerned, arrangements for access to premises and data, processes for the clearance of staff, data retention periods, etc.).

FOR APPLICATION DEVELOPERS

Some assistants – especially those most popular with the general public – rely on platform-based approaches to host third-party applications. In practice, it is generally necessary to respect the development constraints imposed by the designer of the assistant. However, good development practices can also be observed.

Implementing the principles of privacy by design and accountability

Application developers are often offered software development kits (SDKs). As such, they need to approach the life cycle of the data within a specific framework: What data movements are there? Where does responsibility lie? What are the good practices to be implemented? The chapter "Control your libraries and SDKs" in the CNIL "Developer's Guide" covers the technical points to be checked (see box page 75).

Best practices

- Be transparent and explain the different processing steps from the collection phase by the assistant designer to the application's response, specifying which data is accessed by each actor, why and for how long.
- Collect only the data necessary to complete the application.
- Verify that data collection for the application does not trigger further data collection by the developer or associated third parties.
- Regularly check SDK and API functions and the data collected through these channels.
- Control and secure the personal data transmitted from the application to the user through his/her voice assistant, taking special precautions for personal data highly revealing about the user's private life (energy consumption data, bank balances, health data etc.).
- Clearly define the terms of contracts and commitments regarding privacy issues related to the use of resources made available by the assistant designer and do not enter into generic contracts that do not take into account the specificities of the application's needs.
- Clearly specify the chain of responsibility involving the assistant designer and the application developer. Within the scope of responsibility of the company developing its application, implement the obligations of the GDPR, in particular informing individuals, the exercise of their rights and data security.
- Take full advantage of the possibilities enabled by the SDK's designer to deliver clear information and propose appropriate authentication mechanisms during the initial configuration.

FOCUS ON...

GDPR developer's guide

In order to assist developers in the compliance of web or application projects, the CNIL has drawn up a best practice guide for open source development[149][150]. It offers advice and good practices, and thus provides keys to understanding the GDPR that are useful for all stakeholders, regardless of the size of their structure.

This guide is divided into 16 thematic sheets that cover most of the needs of developers to accompany them at each stage of their project, from development preparation to audience measurement. These good practices, which are therefore not intended to cover all the requirements of the regulations nor to be prescriptive, provide a first level of measures to take into account privacy protection issues in IT developments that are intended to be applied to all projects processing personal data.

[149] - CNIL, GDPR developer's guide, https://www.cnil.fr/en/gdpr-developers-guide
[150] - https://github.com/LINCnil/GDPR-Developer-Guide

FOR VOICE ASSISTANT INTEGRATORS

Although they are software products, voice assistants are intended to be embodied in physical equipment. These connected objects can be very diverse: smartphone, vehicle, household appliance, living room speaker, children's toy, etc.

In some cases, designers and integrators of voice assistants may be the same entity, but this is not necessarily the case. While voice assistant designers focus on the software aspects (while providing guidance in terms of required specifications), integrators focus on the hardware constraints. It should be noted that there are many cases and contexts of use and different target audiences depending on the applications. The advice outlined here is therefore generic, but can be added to depending on the modalities of use. In all cases, particular attention should be paid to the way in which individuals are informed.

Establishing transparency as a basis for trust

As for any processing of personal data, the GDPR requires that data subjects be informed of the processing operations carried out by voice assistants. Even more than for online services, the use of a voice assistant requires users to trust a device that reveals little about how it works and whose control means are not as integrated as those used on a computer or smartphone.

Best practices

- Verify that the information and transparency conditions provided by the designer of the assistant are satisfactory to enable the processing of personal data in accordance with the legislation (see the advice for designers of assistants above, page 68).
- Provide the planned information and, where appropriate, adequate additional information.
- If there are plans, in a future software upgrade, to equip an object with a voice assistant, clearly indicate in the equipment specifications whether sound production (loudspeaker), listening (microphone) and computing (processor) capabilities are embedded.
- Specifically inform users when a voice assistant feature is deployed on an equipment that did not initially offer it, and allow them to continue to benefit from fully functional equipment without activating the voice assistant if they wish.

Giving control to users

Another major requirement of the GDPR is the implementation of means to enable individuals to control the uses that are made of their data and to exercise their rights in a simple and effective way. These control and exercise modalities must be adapted to the assistant's voice interface.

Best practices

- Think beforehand about the interest and expectations of integrating a voice assistant in the equipment in question.
- If such a choice is indeed relevant, choose the assistant to be integrated according to the objectives pursued and the need to protect privacy.
- Allow the user to choose whether or not to use the built-in assistant if it is not absolutely necessary for the proposed service, while continuing to benefit from fully functional equipment
- Implement a physical microphone mute button (electrically powered).

Satisfying the security requirement

In the same way as for the designers of voice assistants, the GDPR specifies that the protection of personal data requires appropriate technical and organisational measures to ensure a level of security appropriate to the risks. The implementation of any processing of personal data therefore implies an obligation of security. In addition to the generic measures that can be found in the guide "Security of Personal Data"[151] and in the "Developer's Guide"[152] (presented in the box on page 75), good practices specific to voice assistants can be specified.

Best practices

- Deploy voice assistants on updated and properly secured equipment (see for example the box on connected toys, below).
- Opt for assistants whose operation is controllable, i.e. for which it is possible to act on all the technical parameters and the selection of functionalities.
- Avoid assistants that may pass on data to a third party without knowing the conditions under which the data is processed by the third party.
- Avoid assistants operated by actors who reuse data for their own account or ensure contractual regulation of processing operations carried out by the designer of the assistant.
- Carry out a Data Protection Impact Assessment and regularly update it to ensure that the technical and organisational measures taken are in line with the risks that data processing poses to individuals (for more details on the DPIA, see box on page 56).

FOCUS ON...

Connected toys not always safe

In 2017, the CNIL carried out verification missions on two connected toys. These toys, equipped with a microphone and a speaker, answer children's questions on various subjects such as fairies and dinosaurs. The answer is retrieved from the Internet and given to the child through these objects.

The checks carried out revealed that the company marketing these toys collects through them a multitude of personal information on children and their entourage, including their voices and the content of conversations exchanged. Moreover, it has been found that the lack of security of toys allows anyone with a device equipped with a Bluetooth communication system to connect to it, without the knowledge of the children and the adults around them, and thus to have access to discussions exchanged within a family or among friends.

In light of these elements, the CNIL's President considered that the processing operations carried out did not comply with the French Data Protection Act (loi Informatique et Libertés) because of the failure to respect the privacy of individuals and the lack of information for the data subjects, and therefore decided to give formal notice to the data controller to adopt corrective measures within two months. This formal notice was made public in December 2017.

[151] - CNIL, Security of Personal Data, 2018 edition https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle_gb_web.pdf
[152] - CNIL, GDPR developer's guide, https://www.cnil.fr/en/gdpr-developers-guide
[153] - CNIL, [Infographie] Il était une fois l'ours connecté mal sécurisé, 2017, https://www.cnil.fr/fr/infographie-il-etait-une-fois-lours-connecte-mal-securise

FOR ORGANISATIONS WISHING TO DEPLOY VOICE ASSISTANTS

As the strategies of different voice assistant designers seem to indicate, there is a strong tendency towards the deployment of these assistants in shared environments. There are many initiatives along such lines: partnerships with hotel chains, standard integration in vehicles, some of which will be rented, product line for the corporate world, etc.

In any case, it should be noted that there are many cases and contexts of use and different target audiences depending on where these technologies are implemented. As developed in Chapter II.2 *What issues for voice assistants?*, deployment in open locations or places of transit raises a great many questions that need to be answered before implementation. This is particularly the case for public places, not dealt with here, for which a specific legal framework is necessary. The advice outlined here is therefore generic, but can be added to depending on the uses. A professional use does not carry the same risks and obligations as a more recreational one, which is still different from that which would be made by dependent people. In all cases, particular attention should be paid to the way in which individuals are informed.

Establishing transparency as a basis for trust

As for any processing of personal data, the GDPR lays down a duty to inform data subjects of processing operations carried out by voice assistants. Even more than for online services, the use of a voice assistant requires users to trust a device that reveals little about how it works and whose control means are not as integrated as those used on a computer or smartphone. In the case of a professional environment, the information must particularly be adapted to the context in which the assistant is being set up. Thus, particular constraints may apply, such as consultation with employee representative bodies or, more broadly, labour law, or the consideration of specific audiences.

Best practices

- Inform all persons likely to see their interactions recorded by the device.
- Where appropriate, provide for a method of obtaining consent from persons and an alternative method of operation, necessary for free consent.
- When embedded in a dedicated device, position the voice assistant in a place where it will be prominently displayed and visible to all.
- Determine whether certain categories of vulnerable persons (elderly, dependent, children, etc.) are likely to be affected and take the necessary measures (see box opposite).

FOCUS ON...

The implementation of a voice assistant for dependent persons

In the case of assistive devices for persons with decreasing independence, the chain of responsibility for voice assistants involves, in addition to users, designers and developers of third-party applications, a third-party helper. Whether it is a family member, a home support worker, a social worker, an independent service provider, or a representative of the equipment manufacturer, his/her intervention may be necessary to set up and configure the device. Depending on the case, some of these stakeholders may be linked to the end-user by a contract for the sale of equipment or subscription to the service, installation and maintenance services, etc. Others, including family members, may be materially involved in the setting up or operation of the voice assistant, without their role being formally defined. In these circumstances, the qualification of the status of these persons under the Regulation – as data subject, controller, processor, etc. – can only be done on a case-by-case basis. In extreme cases of loss of independence, the very capacity of data subjects to perform legal acts – including the act of consenting to the processing of their data – could be problematic. In light of these considerations, consent does not therefore appear, in general, to be an appropriate legal basis.

The data that may be collected is the same as that which is used conventionally: identity data, authentication data, contact information data, etc. In the case of the intervention of a third-party carer (a relative, a home helper or a service provider) specific guarantees must be put in place. This is in order to limit the risks of data breach, invasion of privacy or identity theft, since certain data, in particular authentication data, is intended to remain confidential and known only to the data subject.

In addition, it should be noted that certain functionalities targeting people with decreasing independence may require the processing of sensitive data (for example, a reminder for taking medication). Similarly, the way in which voice assistants are used can reveal some of the vulnerabilities of users (e.g. amplification of sound during a telephone conversation, making an emergency call, or inconsistent voice commands). Finally, metadata related to the use of the device may provide, in case of access by an unauthorised third party, indications on the physical activity (e.g. whether or not the user is present at home) of the user. Unauthorised access to such data is likely to generate risks that are all the more important since users of these devices can often live in relative isolation and/or be in a situation of decreasing independence (and, therefore, of physical or cognitive fragility). Therefore, when designing and manufacturing these devices, great care must be taken to ensure their safety. As such, an impact assessment relating to the protection of personal data may then be necessary (see box on page 56).

Giving control to users

Another major requirement of the GDPR is the implementation of means to enable individuals to control the uses that are made of their data and to exercise their rights in a simple and effective way. These control and exercise modalities must be adapted to the assistant's voice interface.

Best practices

- Opt for devices equipped with a physical microphone mute button.
- Consider a means of activation that is less uncertain than wake word detection (e.g. by activating a physical button).
- Leave the possibility of muting the microphone to users.
- Choose the assistant to deploy according to its features and specifications. For example:
  - How is the data managed?
  - Is any data reused?
  - Implementation of local/remote data processing?
- Ensure that users have the means to exercise their rights to their data (information, consultation, access, deletion, objection), for example by making the activation of the device conditional on the provision of a means of contact (email address for example).
- Opt for voice assistants that offer a private navigation mode for actions that do not require authentication, thus allowing a user to interact without an account being associated, nor any trace of these interactions being kept.
- Configure the assistant to reset itself at short notice so that no data is retained beyond the intended interaction, especially in transit locations.

Satisfying the security requirement

In the same way as for the designers of voice assistants, the GDPR specifies that the protection of personal data requires appropriate technical and organisational measures to ensure a level of security appropriate to the risks. The implementation of any processing of personal data therefore implies an obligation of security. In addition to the generic measures that can be found in the guide "Security of Personal Data"[153] and in the "Developer's Guide"[154] (presented in the box on page 75), good practices specific to voice assistants can be specified. In particular, if the assistant is made available in a place open to the public or on a network accessible to a large number of users, securing its use requires additional measures.

Best practices

- Carry out a Data Protection Impact Assessment (DPIA) and update it regularly to ensure that the technical and organisational measures taken are in line with the risks that data processing poses to individuals (see box on page 56).
- Deploy voice assistants on updated and properly secured equipment (see the box on connected toys on page 77).
- Carefully choose the services that can be controlled by the voice assistant, identify those potentially at risk and strictly supervise the administration of the assistant.
- Be vigilant in installing and accessing only legitimate applications, as hackers can create malware to collect data or enter the organisation's information system.
- In the case of deployment in a professional environment, take into account the various risks that may weigh on the organisation from the idea stage: risks regarding privacy, cybersecurity or the confidentiality of certain sensitive or strategic information for the organisation, etc., and take measures accordingly.

[153] - CNIL, *Security of Personal Data*, 2018 edition https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle_gb_web.pdf

[154] - CNIL, *GDPR developer's guide*, https://www.cnil.fr/en/gdpr-developers-guide

Respecting the rights of employees

The decision to deploy a voice assistant in a professional environment may be motivated by the desire to facilitate employees' day-to-day, improve the work tools at their disposal, etc. While such tools may appear legitimate, they must not lead to employees being placed under constant and permanent surveillance. The introduction of such technical equipment should therefore be closely supervised.

According to the case law of the French Court of Cassation in social matters, employees must be informed of the periods during which they are likely to be listened to or recorded. Even if this is not the purpose of installing a voice assistant in a workspace, the risk of misappropriation of recordings exists.

Best practices

- Inform employees individually (email, appendix to the employment contract if necessary, etc.) and collectively (via information and consultation, where appropriate, of staff representative bodies) prior to the deployment of one or more voice assistants.
- In particular, the information must specify where these devices can be deployed (meeting room, employee's office, etc.), who can access them, for what purposes, for how long, and what rights employees have in this respect.
- Clearly define the chain of responsibility involving the employer, the assistant designer and the application developer.
- When embedded in a dedicated device, position the voice assistant in a place where it will be prominently displayed and visible to all.
- Supervise the use of such devices (places where they can be deployed, the conditions for their start-up and shutdown, the process of consultation by the employer of the data collected or generated by the devices, possible sanctions in the event of non-compliance with the instructions, etc.). These details may be included in the company's internal rules and regulations or IT charter.
- Provide for the deletion of personal data, user accounts, etc., for employees whose employment contract would be terminated.

.

FOR END-USERS

Terms of activation and information, available services and uses, security measures... choosing to use a voice assistant is not insignificant. It is important to be aware of the challenges posed by these devices. Five points of vigilance should be noted for users.

Ensuring the confidentiality of exchanges

Voice assistants being in permanent standby mode, they can activate and inadvertently record a conversation as soon as they assume to have detected a wake word. Once recorded, the interactions might be listened to by persons, employees or service providers of the company providing the voice assistant, in order to improve the various algorithms implemented (wake word detection, automatic speech transcription, language comprehension, etc.). Choosing to place such a device at the heart of one's home or vehicle therefore implies responsibilities towards the various persons whose personal data may be processed.

Best practices

On the choice of the device to be used

- Give preference to the use of devices performing local data processing over those performing remote processing.
- Opt for devices equipped with a physical microphone mute button.
- Favour devices that allow the activation of listening by manual pressure on the device rather than by a wake word, which will give you more control over its activation periods. Otherwise, give preference to devices that signal the start and end of recording periods with an audible signal and activate them when installing the voice assistant.

On the use of the device

- If you don't want people listening to your conversations and your device allows it, disable the analysis of your interactions for product improvement purposes.
- If you do not wish to share technical data, disable the analysis of technical data for product improvement purposes.
- Mute the microphone/turn off the device when you do not wish to be listened to by the assistant. Please note that some devices do not have an on/off button and must therefore be unplugged.
- Notify third parties (guests, household staff, etc.) of the potential recording of conversations, or mute the microphone/turn off the device.
- Conversely, if you stay temporarily in a place where a voice assistant is present, ask the owner to disable it or unplug it if you do not wish to be recorded.
- When embedded in a dedicated device, position the voice assistant in a place where it will be prominently displayed and visible to all.
- Regularly check the history of recorded data in the user account and delete confidential data.

Monetising your personal data

Primarily intended for use in the home (or its extension, the personal vehicle) to control connected objects and entertainment services, devices equipped with a voice assistant are at the heart of home life. In many cases, the user's different interactions with the assistant feed into a profile related to the latter. Lifestyle habits (time of getting up and going to bed), heating control, cultural tastes, past purchases, interests, etc., all this information can then be used for advertising targeting.

Best practices

- Be aware that what you say in front of the device can be used to develop your advertising profile. Certain assistant designers allow you to view the ad segments in which you have been categorised and remove that categorisation.
- Opt for devices that do not require the creation of a user account for their use.
  - When the device requires the use of a user account, or when certain features make the use of an account necessary, evaluate whether it is preferable to link an existing account or, on the contrary, to create a dedicated account.
  - When the use of an individual account is required for certain features, keep in mind that anyone with access to the assistant will be able to use them once it is installed, unless you set up authentication measures.
- If the assistant allows this functionality, opt for the use of a "non-linked" mode (private browsing) to its accounts when the connection is not necessary to process and execute the order placed.
- Only connect to the assistant services that are genuinely useful, while considering the privacy risks of sharing private data or sensitive functionality.
- Regularly check which services are connected to the assistant, and disable seldom or unused services.
- Do not hesitate to contact the support services if you have any questions and to exercise your rights with them (for example the right of access), and, if necessary, the CNIL.

Remember that there is no screen

While a companion screen is often necessary to configure the assistant, the ambition of voice assistants is to offer interactions that do not rely primarily on a visual medium. However, without a screen, it is sometimes difficult to get an overview of recordings, to judge the relevance of the suggestions made, to find out more or to access answers from other sources.

Best practices

- Opt for devices that allow settings management and data erasure via the voice interface in addition to the option via the companion screen or user account.
- Regularly consult the assistant's management dashboard to customise its features according to your needs. For example, define the default search engine or information source used.
- Do not hesitate to use the assistant's features to program reminders of the tips presented in this chapter!

Supervising use by children

Initially an object of curiosity, voice assistants can quickly become a digital interface particularly appreciated by children for its (relative) ease of use. While there is no doubt that a computer or smartphone should not be left in the hands of a young child without parental supervision, it is essential to note that the same is true for voice interfaces.

Best practices

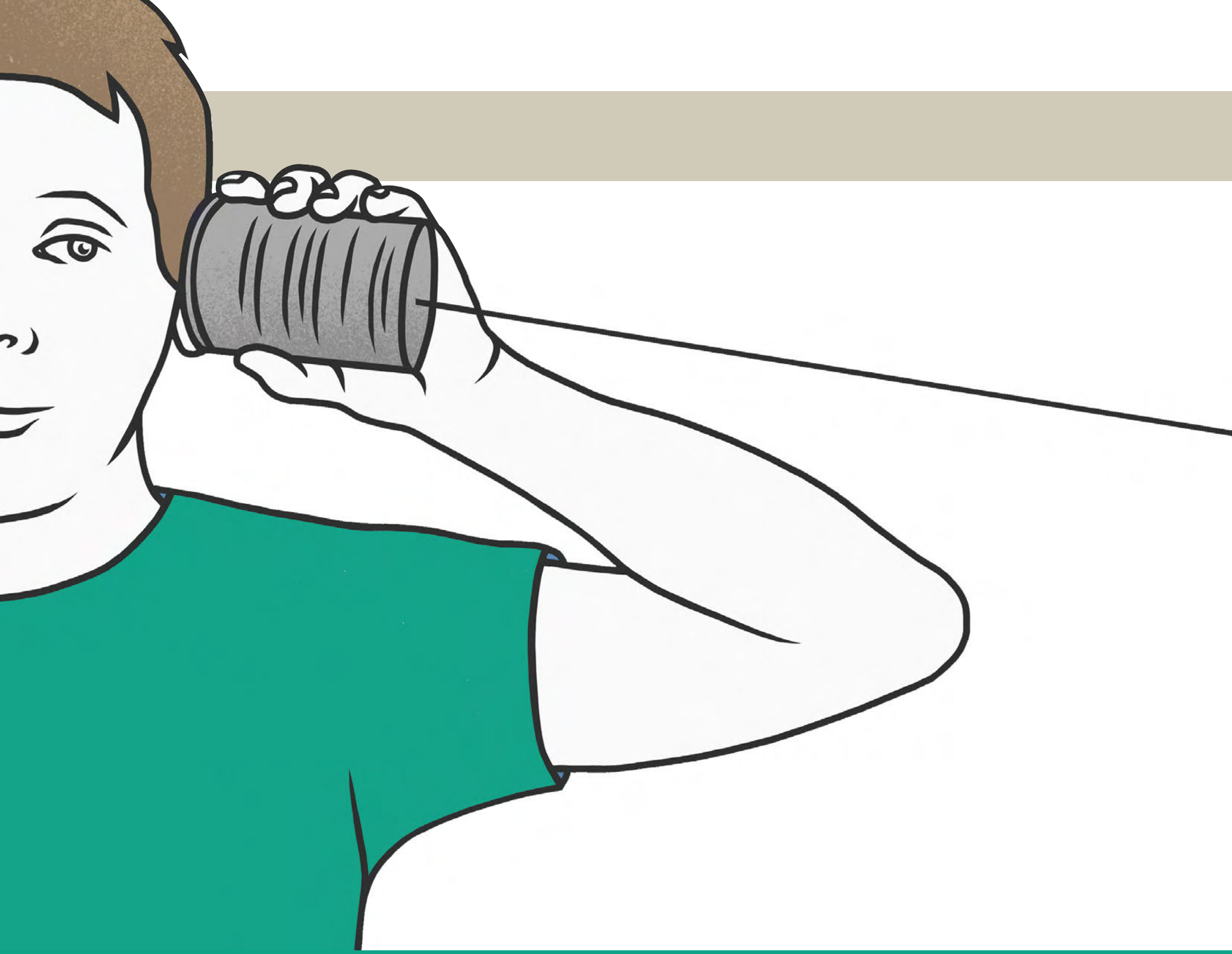
- Explain in a clearly instructive way how a voice assistant works and show the simple settings (deactivation button for example).
- Avoid deploying these devices in areas reserved for children (bedroom, playroom, etc.).
- To supervise children's interactions with the device: stay in the room when they're using it, turn it off when you're not with them.
- Make sure it is set by default to filter information for children.
- If a history is recorded, consult usage statistics and, if applicable, past interactions, while respecting the child's privacy.
- Delete this history on a regular basis.

Preventing hacking risks

Depending on the choices configured by the user, different services can be accessed by a voice assistant. However, it does not always offer an authentication possibility to ensure the legitimacy of the person placing the order. It should therefore be borne in mind that the voice assistant may, if it is connected to many services (e.g. home automation or banking), represent a vulnerability in the household's information system.

Best practices

- As with any connected object, avoid products whose origin and designer are not recognised or for which it is not possible to easily identify the controller and a contact point, ideally in French.
- Carefully choose the services that can be controlled by your voice assistant and avoid those at risk (door opening, lock, starting a vehicle, etc.).
- Be careful to install and access only legitimate applications, as hackers can create malware to collect user data (account or credit card number, password, address, contact, etc.).
- Set up the security of the device or certain sensitive applications, through two-factor authentication (e.g. via a verification code sent by email or text message) if the device allows it.
- Carefully choose the activation in the assistant of services related to your accounts (emails, calendar, bank account, calls, etc.) that would be accessible by anyone in the same room.
- Secure the network (especially Wi-Fi) to which the assistant is connected.



Commission Nationale de l'Informatique et des Libertés
3 place de Fontenoy
TSA 80715
75334 PARIS CEDEX 07
Tél. +33 (0)1 53 73 22 22
www.cnil.fr

November 2020

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS