MINISTÈRE DE LA CULTURE

Liberté
Égalité
Fraternité

HIGHER COUNCIL FOR LITERARY AND ARTISTIC PROPERTY

MISSION REPORT

VOICE ASSISTANTS
AND OTHER CONVERSATIONAL AGENTS

Mission Chair:
Célia Zolynski
Qualified person at CSPLA
University of Paris 1 Panthéon-Sorbonne

In collaboration with:
Karine Favro, University of Haute-Alsace
Serena Villata, University of Côte d'Azur, CNRS, Inria

December 2022 - Final version

Software applications offering oral dialogue capabilities with users, voice assistants have experienced exponential growth since their entry into the market. While they are not (yet) a key access point for the cultural sector, they are a particularly fast-growing application in this area. Voice assistants are now the developing interface for users to access the web, connected devices and IoT services in a more natural way, especially with recent advances in machine learning and automatic Natural Language Processing (NLP) methods. This is partly because voice control offers many advantages in building human-machine interaction, making voice a new interface, simplifying interactions and promoting user engagement. At the same time, voice includes new challenges (correct interpretation of the user's request, referencing of proposed offers, access to information for the user, processing of the most intimate data) that can reduce the choice of users. And these issues are amplified when it comes to voice assistants, given the so-called single response mechanism, at the risk of reducing the diversity of cultural offers that the user could access via this interface. This also results in important competition issues, notably due to the creation of large databases by a few market actors and questions about access to the usage data collected, as well as the limited possibility of parametrisation and the presence of applications integrated by default in the voice assistant universe.

These new forms of interaction through natural language therefore imply a renewed reflection on the qualification and the legal regime of these systems and their effects on the diffusion and diversity of cultural content. In this respect, current practices invite us to focus on voice assistants. However, the evolution of these technologies should also raise questions insofar as it heralds new human-machine interfaces that will lead to increased user engagement.

To this end, this report, designed within the framework of an interdisciplinary approach, proposes an explanation of the technical functioning of voice assistants and, more generally, of conversational agents as well as their universe. This clarification is combined with the underlying legal questions concerning the cultural sector (I). It also considers, more specifically, the issues of market access and the responses provided by the recently adopted texts (II) and then analyses the conditions for guaranteeing diversity, which implies defining the role that the user must play (III).

As technological solutions, voice assistants can be considered from two different, yet complementary, points of view to get an overview of their impact and the issues that arise from them. First of all, the voice assistant can be studied as an algorithmic system that can process a request formulated by the human voice and answer it by a synthetic voice. The complexity of this architecture is linked to the massive collection of highly heterogeneous data which, in conjunction with increasingly powerful machine learning algorithms, allows the relationship between the human and the voice assistant

to be locked down to the exclusion of third-party user companies, such as application developers and content publishers. Secondly, the voice assistant should be seen as an interface to many related contents and services, offered either by the voice assistant provider or by third party companies. It also constitutes an additional functionality for accessing an ecosystem, and can even become the entry point for accessing a set of digital services offered by its provider, characterizing a new form of intermediation that is at the heart of recent reforms of European Union law regarding digital markets and services.

Concerning market access, the network effect is particularly developed in the voice assistant market in that it influences negotiations between the different actors in a market deployed as an ecosystem. The structuring actors occupy strategic positions in the form of a powerful oligopoly, but above all, they are able to make user companies adhere to their environment. These actors have the ability to filter content, which is due to the "result engine" nature of the voice assistant. This helps to significantly increase the bargaining power of service distributors (such as manufacturers of connected speakers and other connected objects that extend the ecosystem or providers of voice assistants). Similarly, manufacturers of technical solutions may have significant bargaining power, particularly with respect to content publishers. The cost of developing voice applications, the cost of accessing the catalogue, and the technical effort to integrate the system are imposed on them, while promising mutual and effective collaboration. Finally, the accumulation of powerful intermediaries between content publishers and users raises questions about the capture and distribution of value for publishers who are at the end of the contractual chain. And the user is not protected from this effect. He has priority access to services and products highlighted by the brand that can strongly identify the voice assistant. They are poorly informed, sometimes subjected to misleading interfaces, pre-installed applications by default, and a lack of visibility of certain content, which can hinder their access to culture and lock them into their choices. The single answer implies a real restriction of the exposure of works and services, and even of access to the market.

In order to grasp all the issues at stake in the voice assistant market, we examine more specifically the key points that will allow non-discriminatory and equitable access to the market so as to guarantee the user's freedom of choice, particularly when accessing cultural content: self-preference, interoperability of systems and applications, and access to technical data and user data. The Digital Markets Act (DMA) has addressed these issues through its cross-sectoral approach, which includes families of obligations for gatekeepers.

These new forms of human-machine interaction also have important consequences for the diffusion and diversity of cultural content. This implies thinking about the instruments to guarantee diversity within these ecosystems. In order for the user to exercise freedom of choice, it is then appropriate to question the practices amplified or specific to voice assistants as soon as they alter the access to content for the user,

particularly because of the mechanism of the single answer and its risks of confinement. The operation of these devices also relies on a significant collection of personal data, through the simple recording of voice requests as well as through the analysis of the content of these requests, which may compromise the user's freedom of choice.

This implies acting on different levers, according to a three-step approach. First, the restrictions on the market must be lifted by questioning the self-preference practices of voice assistant operators favouring their own services, especially when the voice assistant responds to the user's voice query by highlighting its related services. Self-preference can then be viewed from the perspective of restrictive market access and anti-competitive practices. Beyond guaranteed access to a plurality of content, freedom of choice also requires ensuring the diversity of this content. This leads us to question the technical integration of existing applications to guarantee diversity, that can not only guarantee access to several contents, which is what pluralism comprises, but can also to guarantee access to different contents. This supply-side approach to diversity must be complemented by a demand-side approach to diversity. The role of users should be thought of differently, by giving them the means to be informed but also to act on access to content, in particular by giving them a right to parametrisation to widen their choice of content in the ecosystem, but also to choose this ecosystem by removing the barriers to exit, which implies facilitating the conditions of unsubscribing to the service and guaranteeing them a right to portability of their data. This amounts to thinking about the mechanisms of co-regulation of diversity by supply and demand.

Table of contents

Introduction

General presentation. 2018 is a landmark year for voice assistants, which signals an end, that of smartphones as designed, and a beginning, that of connected object onboarding[1]. Users are saturated with screens, whether they are smartphones or computers, fuelled by talk of public health issues. Especially since the health crisis, this has produced a commitment to other modes of consumption and interaction in favour of a renewed interest in the audio channel and the [2]rise of connected speakers generally integrating the voice assistants of Google and Amazon[3]. Consumer voice assistants were first accessible on smartphones and other touch tablets, but they are now developing on connected environments to build the smart home of tomorrow. The assistant will then become "*a universal remote control*", which could disqualify "*equipment that is not compatible with its software layer*"[4]. The objective of the next few years will be to design invisible interfaces worn by users in order to accompany them in their travels, possibly to virtual worlds. So much so that the saturation threshold linked to the time spent on screens could quickly be exceeded by listening to new media designed to entertain and then assist the user on a daily basis[5].

Since their entry on the market, the growth of voice assistants has been exponential. Now, " *voice assistants represent the fastest developing interface for users to access the web, to use and control smart devices and access consumer IoT services* "[6]. Their use is now intended to be more ambitious. Recent advances in machine learning and Natural Language Processing (NLP) methods, combined with the availability of computing power, have made new algorithms available to developers, created to make interaction more natural. Today, conversational agents can be considered as a "*mature and deployed technology, but it remains perfectible*"[7], waiting for a new generation.

Voice assistants, chatbots, conversational agents: a variety of terminology and techniques. Chatbot, voice assistant, virtual assistant and conversational agent are all terms used to designate the same thing at first glance, namely the fact of dialoguing with a machine. This means that the translation of the term "chatbot" clearly expresses an action that characterizes a human/machine interaction present when referring to the voice assistant, virtual assistant or even conversational agent. Are these generic terms used indiscriminately by the designers and authorities who take up this innovation? The question is worth asking because the answer does not seem to be fixed.

---

[1] HADOPI/CSA, *Assistants vocaux et enceintes connectées. L'impact de la voix sur l'offre et les usages culturels et médias*, May 2019, p. 6.
[2] O. Gouliaéva, E. Dosquet, Y. Moysan, *La révolution des assistants vocaux,* DUNOD, 2020, p.11 and following.
[3] HADOPI/CSA, Study supra.
[4] France Stratégie, *Le monde de l'internet des objets : des dynamiques à maîtriser*, February 2022, p. 107.
[5] O. Gouliaéva, É. Dosquet, Y. Moysan, supra.
[6] European Commission, *Final Report - Sector inquiry into consumer Internet of Things, {COM(2022) 19 final},* Jan. 20, 2022, point 28 - On the evolution of the market, see HADOPI/CSA, study supra.
[7] France Stratégie, Report supra, p. 87.

While some consider it so, the reality seems more nuanced as the terminology depends on the level of conversation. In other words, the criterion that could distinguish a voice assistant from a chatbot proceeds from the construction of an exchange allowing the users to receive an answer not only to the most obvious questions they will ask the chatbot, but also to the underlying questions they ask themselves. The chatbot can thus analyse the user's intentions from a simple action such as consulting bank accounts, and can even lead him to understand his intentions that he did not always premeditate, such as the reason why he consults his bank accounts. The process is identical with the weather, or even music, etc. When the conversational agent is capable of this level of conversation and of maintaining the conversation, it is called a chatbot. The chatbot can go further by reinforcing the personalization of the conversation and the day-to-day support of the user. Not only are the answers personalized, but they are also intended to blend in with the user's everyday behaviour. The approach is more contextual, requiring context and behavioural data. From then on, we readily move from the simple assistant (offering a FAQ / chat type response for online discussion) in that it provides a contextualized response by facilitating a service rendered to the user, to the chatbot which is a "*digital butler at our service, listening to us*"[8] in that it integrates personalized responses to complex requests and automated processes to perform certain actions.

In reality, the virtual (voice) assistant represents the first generation of conversational agent showing a human/machine interaction that is still embryonic, or even technically limited (e.g., by the voice interface, interoperability, market access conditions) in the provision of content, which raises questions about their capacity to inform or entertain in terms of respect for pluralism. The chatbot can also characterize a vocal interaction with the user, but it could be even more advanced, going as far as cloning the user's voice in vocal synthesis. Consequently, it is a question of providing the chatbot with the keys to understanding human emotions, which immediately constitutes a risk of confining the user in his own universe.

Voice assistants, chatbots, conversational agents: definitions. It is therefore appropriate to clarify the terminology and present the definitions to be used to analyse the issues related to the deployment of voice assistants and other conversational agents that are the subject of this report. In general terms, conversational agents or dialogue systems[9] are software that communicate with users in natural language (through text, speech, or both) and fall into two categories. The first category includes virtual assistants (VAs) that use conversation with users to help them accomplish specific tasks. Some virtual assistants are able to interpret human speech and respond with synthetic voices. These virtual assistants are called "voice assistants". Users can ask questions to their general-purpose voice assistants, control home automation devices and media playback by voice, and handle other basic tasks such as sending

---

[8] FING, *Hypervoix. Promesses et illusions des nouvelles interfaces humain machine*, January 2019, p. 9.
[9] D. Jurafsky, J. H. Martin, "Speech and Language Processing," *Prentice Hall*, 2021, https://web.stanford.edu/~jurafsky/slp3/

email using verbal commands. Siri, Alexa, Google Now/Home and Cortana are examples of general-purpose voice assistants. Voice assistant technology is a must for connected speakers, for which it is the only mode of interaction[10]. The second category includes chatbots, which are systems designed for extended conversations that mimic the unstructured conversations or "chats" characteristic of human-human interaction, primarily for entertainment purposes, but also for practical purposes, such as making virtual assistants more natural. This is the case, for example, of BlenderBot from Meta or the XiaoIce system from Microsoft. Some chatbots have a more developed personality than others; they are then called social chatbots.

In the following, virtual assistants will be referred to as task-oriented conversational agents and chatbots will be referred to as systems that are able to engage in a prolonged conversation with the user, without a specific task to perform. When the virtual assistant interacts with the user by voice, the term "voice assistant" will be preferred to designate it.

Distinction made between general or general-purpose voice assistants and specialized voice assistants. The European Commission has been looking at voice assistants as part of its sector inquiry into consumer Internet of Things[11]. As such, the proposed definition is as follows: "*a voice-activated piece of software that can process voice commands and return relevant information or perform certain functions as requested by the users* ". Indeed, the voice assistant acts both as an intermediation platform for voice applications and as a user interface. This interface is now developing rapidly because of the ease with which the user can make a request with voice to access the web, use and "control" smart devices. Voice interaction is useful in certain constrained environments such as the car, or the home, as a complement to another activity. The voice assistants that dominate the European market are general-purpose, in that they allow the user to access multiple functionalities identified by the Commission. They are distinct from (or absorb as the case may be) specialized voice assistants provided by IoT service providers or home automation device manufacturers. They can also be configured as simple voice control interfaces.

General-purpose assistants, according to the distinction made by the Commission[12], can play music and videos, listen to the radio, podcasts, news or audio books. These voice assistants are combined with other smart devices to make it easier to listen to or view responses to requests made by the user, such as headphones, speakers or TVs. However, the general-purpose assistant can also provide information that helps the users' daily life, such as the weather, a route, a recipe, gardening tips, music listening, a podcast, etc. Finally, it can participate in the daily routine of the user executing his schedule, such as reminders of events in the calendar, booking meals, managing thermostats in the house, etc. The general-purpose assistant as defined by

---

[10] HADOPI / CSA, Study supra.
[11] European Commission, *Final Report - Sector inquiry,* supra.
[12] European Commission, *Final Report - Sector inquiry*, supra, point 30.

the European Commission is interesting in that it characterizes a real system, or even ecosystem, within the ecosystem of Amazon (Alexa), Samsung (Bixby), Google (Google Assistant) and Apple (Siri). The voice assistant not only allows access to the services offered by the ecosystem from which it emanates, but can also use the third-party applications available in order to provide an answer adapted to the users' request. The voice assistant can, in fact, perform different tasks such as capturing and playing back sound, automatic transcription of speech (using speech-to-text conversion), or the opposite by generating language (speech synthesis that allows text to be translated into speech), automatic language processing, dialogue strategies, access to ontologies (structured data sets linked to a defined domain) and external knowledge sources. The voice assistant is not in itself intelligent, but is integrated with intelligent systems, so the two can be confused.

Other assistants are considered specialized[13] due to their functionality being limited to the services offered by the IoT service providers that bring them to market. This is the case for Orange's voice assistant, Djingo, which authorizes access to Orange services, but also for OK Freebox, OK SFR, Microsoft's Cortana, etc. Specialized assistants, limited by their audiovisual or home automation environment, are generally available in a single language, which is not the case for general-purpose assistants that range from 8 to 30 languages[14]. In principle, there is no interaction between general-purpose voice assistants and specialized voice assistants.

The terminology used in the texts. The guidelines for applying the GDPR to "virtual voice assistants" formally adopt the correct name for this form of voice interaction, which is defined therein as a "*service that understands voice commands and executes them or mediates them with other computer systems if necessary*"[15]. Since the purpose of these guidelines is to provide a procedure for the use of voice as personal data, the focus is on voice control, which is the core element of the system. The Digital Markets Act[16] includes in its scope the more generic category of virtual assistant. Article 2 point 12 of the Regulation defines a "virtual assistant" as "*software that can process requests, tasks or questions, including those based on audio, visual or written input, gestures or movements, and which, on the basis of these requests, tasks or questions, provides access to other services or controls physical connected devices*". The approach taken in the text is more global and forward-looking, not limited to the voice market, but covering all market access issues, including interoperability, self-preference, contractual aspects, etc. This definition is also used by the proposed Data Act[17]. These texts thus capture the voice/virtual assistant in its intermediation function in that it gives access to services provided by the platform and the user companies.

---

[13] European Commission, *Final Report* - Sector inquiry, supra, point 31.
[14] European Commission, *Final Report* - Sector inquiry supra, point 32.
[15] EDPB, *Guidelines on Virtual Voice Assistants*, 02/2021, 7 July 2021, p. 3.
[16] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).
[17] Proposal for a Regulation of the European Parliament and of the Council of 23 February 2022 establishing harmonized rules for fairness in access to and use of data (Data Act), COM(2022) 68 final, Article 2.4.

The voice applications required to operate the voice assistant. We can therefore understand why the study should not be limited to the voice assistant alone, but should include the whole of its "universe", and especially the voice applications that are required to respond to users' requests. These are defined by the European Commission as "*software designed for a specific voice assistant that supports user commands to connect to smart devices, to perform actions or tasks, or to engage in consumer IoT services that consumers access via that voice assistant*". Their naming can be diverse: those written for Amazon Alexa are called "skills", for Google Assistant "actions", for Apple "siri shortcuts". Beyond their name, these applications are specific to each voice assistant and a developer will have to propose a different voice application for each of them. This must be put in perspective with the business model of voice assistant providers, as described by the Digital Markets Inquiry published by the United States House of Representatives in July 2022[18]. Other studies show that voice assistants are in fact a "*continuum of services*" that already exist and "*follow the strategy of the actors who operate them*"[19].

Voice, a new human-machine interface. Voice control offers many advantages in the construction of human-machine interaction, in that it provides ease of use passing from voice to touch and vice versa. It then allows you to perform several activities simultaneously while having your hands busy, to guarantee the immediacy of the interaction, while being only within voice range, and sometimes without being in the perimeter of the voice assistant[20]. In addition, there is "*the naturalness of the interaction that does not require specific learning on the part of the users, the speed of the execution of the command and the extension of the field of application that allows quick access to information*"[21]. The voice becomes a natural interface, enabling the technology to be forgotten. It could thus become the new standard for human-machine interaction. To date, it complements other graphical interfaces (such as computer and smartphone screens) but could eventually replace them to become the primary interface for various computing devices, including IoT devices.

The voice and the unique answer. While voice simplifies interactions, it also brings new challenges, whether it be the correct interpretation of the user's request, the referencing of proposed offers and, more generally, access to information for users of voice assistants, particularly with regard to the processing of their personal data, which is massively collected by this type of interface. It could also reduce user choice, especially in the case of built-in default applications and limited configurability. In addition, the operation of the voice assistant, and particularly the single response,

---

[18] House of Representatives, Subcommittee on antitrust, commercial, and administrative law of the committee on the judiciary of the house of representatives, *Investigation of competition in Digital Markets*, Part I, July 2022, p. 101: " *A Voice assistant platform vendors can monetize its platform by using its ecosystem to drive revenue to complementary lines of business such as e-commerce, search, or entertainment. It can also charge voice-application developers to be the recommended application for a specific command.".*

[19] H. Guillaud, « A qui les interfaces s'adaptent-elles ? », *InternetActu*, 20 January 2019 - for a description of these models, see CNIL, *A votre écoute. Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux,* White Paper No.1, 2019, pp. 24-25.

[20] O. Gouliaéva, É. Dosquet, Y. Moysan, *La révolution des assistants vocaux*, DUNOD, 2020, p.11 and following.

[21] EDPB, *Guidelines on virtual voice assistants*, and already CNIL, White Paper supra, pp. 19-20.

allows for a detailed profile of the user through the accumulation of data targeted by their queries. Real-time machine learning is one of the main features of the voice assistant, which allows it to learn about habits to best adapt its response. The recognized objective results from the consequences of regular interaction to anticipate the user's contextual needs and offer preventive assistance without waiting for his requests.

In the cultural sector, in particular, audio creation has its own characteristics linked to the use of the voice, the work of sound, the narrative, the musical environment, and the particular aesthetic dimension of listening, which are at the origin of new uses. It also can reach new users among the youngest and bring out new talents, especially in the field of podcasts[22]. Beyond that, voice facilitates access to cultural content and refines the relationship with the user by sticking to the listening context but, at the same time, can reduce the user's choice because of the unique response of the voice assistant.

Voice assistants for the cultural sector. Voice assistants are not yet a key access point for the cultural sector, but it is a fast growing application. In music, for example, Sonos recently introduced its own voice assistant, Sonos Voice Control. It is designed specifically for music and has additional privacy guarantees. Sonos Voice Control defaults to the songs and artists that users have preferred in its application during previous interactions. Unlike generalist voice assistants such as Amazon Alexa or Google Assistant, this voice assistant does not upload any voice recordings to the cloud, but processes everything on the device.

In radio, a revolution driven by the voice assistant market is underway. For now, they are used to access content such as podcasts, but their impact could become much greater in the coming years. In the book industry, voice assistants can help users purchase physical books. In addition, the presence on the online stores of the main actors involved (Google, Apple and Amazon) of audio books and digital books that could be read by the connected speaker represents a potential development of these voice assistants. On a more occasional basis, projects have been developed to take advantage of these new means of human-machine interaction, such as augmented reading projects, accompanied by a voice assistant that launches sound accompaniments at the same time as the reading to encourage audience engagement[23], or augmented audio content to envisage a dialogue with the content[24]. Recently, a new connected speaker called Merlin has been offered on the market by Bayard and Radio France. It is a storyteller that comes with 42 preloaded titles (about 4 hours of audio recordings) in its 4GB internal memory. Merlin's capacity is 50 hours

---

[22] Inspection Générale des Affaires Culturelles, *De l'écosystème de l'audio à la demande (« podcasts »): enjeux de souveraineté, de régulation et de soutien à la création audionumérique*, F. Hurard and N. Phoyu-Yedid, October 2020 - in the wake of this report, a study is now being conducted within the Podcast Observatory, under the aegis of ARCOM and the Ministry of Culture (DGMIC), in particular to measure the dynamics of the sector.
[23] In this sense, see the project launched by Disney: https://www.vokode.com/lassistant-vocal-disney-vous-aide-a-devenir-le-meilleur-conteur-dhistoires/
[24] On these different projects, see FING, *Hypervoix*, 2019, pp. 23-25.

in total. More of the 140 additional audio tracks can be added via the Merlin app (available for Android and iOS). This requires the creation of a user account on the official hello-merlin.com website before the content library can be managed and synchronized with the Merlin speaker over Wi-Fi. The selection is varied, with stories written by contemporary authors (Véronique Olmi, Marie Desplechin, Julien Blanc-Gras, etc.), documentaries, songs (in French and English), musical tales and even gymnastics. When it comes to audio quality, Merlin stands out from competing storytellers. It is not, however, a virtual or voice assistant, because the interaction with the speaker is through buttons (a small button to switch it on and 5 buttons arranged in a circle to navigate the menus) and not through speech or text (chat).

Apart from these specific cases, it is difficult for actors in the cultural sector to develop their own voice assistants, which means that they have to form partnerships with suppliers of generalist voice assistants, at the cost of negotiations that are often unbalanced, especially as these offers are generally global. However, these partnerships are essential if producers of cultural content want to reach a large demand. They are then forced to bear the costs of deploying voice applications by each of the voice assistants in order to access this audience.

What prospects for tomorrow? Technical developments allow us to envisage the future deployment of chatbots with human-machine interfaces based on a more natural interaction through speech, which will allow us to adapt the responses of the conversational agent even more finely to the user's request. Simulating a human conversation, the chatbot will be able to identify the emotions felt by its user and to simulate emotions in return - which remain fake - in order to generate an interaction that is always more advanced and a source of engagement by playing on a phenomenon of empathy[25]. This phenomenon could be amplified tomorrow if these chatbots are deployed in virtual universes as some current projects seem to imply[26]. This form of interaction could also support new means of creation, as shown by various artistic projects[27].

Furthermore, special attention must be paid to the development of transformers, which constitute a neural network architecture that can perform complex tasks such as the automatic generation of text or images with reduced training costs[28]. For example, GPT-3 can, from a title, write a paragraph developing the argument contained in the title. It can also be used for creative writing; just ask for it with a short sentence. These

---

[25] See National Digital Ethics Committee (CNPEN), *Ethical issues* of *Conversational agents,* Opinion No.3, November 2021: https://www.ccne-ethique.fr/sites/default/files/2022-05/CNPEN%233-ethical_issues_of_conversational_agents.pdf
[26] See e.g. Meta's Builter Bot project (see Appendix 1).
[27] For example, see J. Walsh's presentation, "Zombie Creativity: the death and resurrection of the author in the age of AI," TESaCO (Academy of Moral and Political Sciences) Project Colloquium "Talking machines: conversation and creation with Artificial Intelligence," 8 November 2021:
https://www.academie-sciences.fr/fr/Colloques-conferences-et-debats/mythes-et-machines.html - also Stephanie Kipkins' project "Not the only one": https://www.stephaniedinkins.com/ntoo.html or Lauren Mc Carthy's project: https://arstechnica.com/gaming/2018/01/artist-transforms-herself-into-a-virtual-assistant-and-obeys-your-commands/
[28] It should also be noted that transformer-based models have begun to have a significant impact outside of natural language. For example, in computer vision, the vision transformer achieves impressive accuracy in image classification tasks by treating small pieces of an image as elements of a sequence, just like a sentence composed of words.

systems are already playing a significant role in the print media industry, as evidenced by the fact that several newspapers including The New York Times, The Associated Press, Reuters, The Washington Post and Yahoo! Sports, use machine learning algorithms to generate content. GPT-3 is an opinion writer for The Guardian and The Associated Press, for example, can now produce 30,000 local news stories a month using AI. These models are also used to build virtual reality and augmented reality environments that are the basis of metaverses.

Issues. Given these multiple issues, these new forms of interaction through language imply a renewed reflection on the qualification and legal regime of these systems**.** The deployment of conversational agents raises issues already considered such as those related to the protection of algorithmic systems[29]. Others are still to be analysed, such as the effects induced by these new forms of human-machine interaction on the diffusion of cultural contents. We observe, for example, that the so-called single response mechanism, the organization of the interface and its ecosystem can reduce the diversity of cultural offers that the user could access via this interface. This also results in important competition issues, notably due to the creation of large databases by a few market actors and questions about access to the usage data collected, as well as the limited possibility of parametrisation and the presence of applications integrated by default in the voice assistant universe.

Scope of the study. Given current practices, the study will focus on voice assistants in order to respond to the challenges resulting from their increasing deployment, while questioning the evolution of these technologies in that it heralds new human-machine interfaces that lead to increased user engagement. The present report, conceived within the framework of an interdisciplinary approach, therefore proposes an explanation of the technical functioning of voice assistants and, more generally, of conversational agents as well as of their universe, combined with the underlying legal questions concerning the cultural sector (I). More specifically, it considers the issues of market access and the responses provided by the recently adopted texts (II) and, in addition, analyses the conditions for guaranteeing diversity, which implies defining the role that the user must play (III).

---

[29] See especially A. Bensamoun and J. Farchy, *Artificial Intelligence and Culture*, final report presented to CSPLA, January 2020.

I - DESCRIPTION OF VOICE ASSISTANTS

Voice assistants are software applications that provide spoken dialogue capabilities with users. These technological solutions can be considered from two different but complementary points of view, in order to draw up an overview of their impact and the issues that arise from them. On the one hand, the voice assistant can be studied as an algorithmic system that can process a request formulated by the human voice and answer it by a synthetic voice. This is the comprehensiveness of this technology, because considering it as an algorithmic system shows the complexity of the upstream architecture of the interaction between the voice assistant and the user. This complexity is related to the massive collection of highly heterogeneous data that, in conjunction with increasingly powerful machine learning algorithms, allows the relationship between the human and the voice assistant to be locked down to the exclusion of third-party user companies, such as application developers and content publishers (A). Also, the voice assistant must be considered as an interface allowing access to numerous content and services that are part of its universe. It also constitutes an additional functionality for accessing an ecosystem, and can even become the entry point for accessing a set of digital services offered by its provider, characterizing a new form of intermediation (B).

A - The voice assistant as an algorithmic system

The voice assistant will allow the human being to speak to a "terminal", i.e. to the machine, to address a verbal request, to which it will respond thanks to a voice recognition and natural language processing system. It is an algorithmic system that receives data related to the human or machine to deduce a way to achieve a series of objectives using learning, reasoning or modelling in order to generate results in the form of content, predictions, recommendations or decisions influencing the real or virtual environment with which they interact. Moreover, the formulation of this response by a synthetic voice represents a significant cost due to complex developments, and is accompanied by the possibility of much greater interpretation errors than a written query. However, it is a very popular solution, especially in areas where safety is a concern, such as in the car. Because it is controlled by voice and leaves the hands free, it is designed to be part of everyday life. In this way, this new human-machine interaction becomes a vehicle for the exposure of cultural content, the development of which has been growing since the first voice assistants were put on the market in 2011.

Conversational agents. Definition(s). Conversational agents or dialogue systems[30] are software that communicate with users in natural language (text, speech, or both) and fall into two categories.

---

[30] D. Jurafsky, J. Martin, "Speech and Language Processing," *Prentice Hall*, 2021.

Firstly, **task-oriented dialogue agents or virtual assistants** use conversation with users to help them to accomplish tasks. More specifically, an **intelligent virtual assistant** is software that can perform tasks or services for an individual based on commands or questions. Some virtual assistants are able to interpret human speech and respond with synthetic voices. These virtual assistants are called **voice assistants**. Users can ask questions of their general-purpose voice assistants, control home automation devices and media playback by voice, and manage other basic tasks such as email and calendars with verbal commands. Existing **generalist/general-purpose voice assistants** such as Apple Siri[31], Amazon Alexa[32], Google Now/Home/Assistant[33] or Microsoft Cortana[34], give directions, control devices, find restaurants or make calls. Virtual assistants can answer questions on company websites and interface with robots. In addition, voice assistant technology is a must for connected speakers, for which it is the only mode of interaction.[35] As highlighted in the EU Consumer Internet of Things Sector Inquiry[36], voice assistants enable communication between hardware (e.g., connected speakers or smartphones) and software components (operating systems such as Google's Android and Apple's iOS for smart mobile devices) and increase their complementarity. In addition, "*these technology platforms also provide centralized access to and control of the products and services they integrate*".

Secondly, **chatbots** are systems designed for extended conversations, which mimic the unstructured conversations or "chats" characteristic of human-human interaction, primarily for entertainment purposes, but also for practical purposes, such as making virtual assistants more natural. Examples include Meta-Facebook's Blender Bot[37], a chatbot based on machine learning methods that can conduct very natural conversations about, for example, music content, or Microsoft's XiaoIce[38], a system that chats with people through a text messages. Some chatbots have a more developed personality than others, and can also offer entertainment and not only assistance for daily tasks; these chatbots are called **social chatbots**.

---

[31] https://www.apple.com/siri/
[32] https://www.amazon.com/b?ie=UTF8&node=21576558011
[33] https://assistant.google.com/
[34] https://www.microsoft.com/en-us/cortana
[35] HADOPI / CSA, *Assistants vocaux et enceintes connectées* supra.
[36] European Commission, *Final report - Sector inquiry into consumer Internet of Things*, 2022, points 164 and following.
[37] K. Shuster, J. Xu, M. Komeili, D. Ju, E. M. Smith, S. Roller, M. Ung, M. Chen, K. Arora, J. Lane, M. Behrooz, W. Ngan, S. Poff, N. Goyal, A. Szlam, Y. Boureau, M. Kambadur, J. Weston, "BlenderBot 3: A 175B parameter, publicly available chatbot that improves its skills and safety over time," 2022: (https://ai.facebook.com/blog/blenderbot-3-a-175b-parameter-publicly-available-chatbot-that-improves-its-skills-and-safety-over-time/)
[38] L. Zhou, J. Gao, Di Li, H.-Y Shum, "The design and implementation of XiaoIce, an empathetic social chatbot", *Computational Linguistics*, 46(1), 2020, pp.53-93.

History of conversational agents

Conversational agents have evolved with the gradual increase in computing capabilities and advances in Natural Language Processing (NLP) tools and techniques. The first implementation of a conversational agent, which relied heavily on hand-encoded linguistic rules, was in 1966 with the development of ELIZA[39]. ELIZA was designed to simulate a Rogerian psychologist, a branch of clinical psychology whose methods consist in making the patient talk by feeding back his statements. If a patient says, "I took a long boat ride," and the psychiatrist responds, "Tell me about boats," it is not assumed that the patient does not know what a boat is, but rather that he or she has a conversational purpose. This conversational agent can communicate with the user through an algorithm that finds a match between the keywords used in the user's request and those present in the rules used to reformulate the input and provide an output, i.e. a response to the user. Nevertheless, the scope of ELIZA's knowledge was limited because it depended on minimal context identification and, in general, the rules defined are not flexible enough to be easily implemented in new domains. A few years after ELIZA, another conversational agent focused on clinical psychology, PARRY[40], was used to study schizophrenia. In addition to the ELIZA-like rules, the PARRY system included a model of its own mental state, with affect variables for the agent's levels of fear and anger; certain topics of conversation could cause PARRY to become angry or suspicious. If PARRY's anger variable is high, he will choose from a set of "hostile" responses. If the input mentions his delusional topic, he will increase the value of his fear variable and then begin to express the sequence of statements related to his delusion. Parry was the first known system to pass the Turing test (in 1972); psychiatrists could not distinguish text transcripts of PARRY interviews from transcripts of interviews with real paranoids.

In the 1980s, the field of conversational agents underwent an important evolution with the use of Artificial Intelligence (AI). Notably, ChatScript marked the beginning of a new era in the technological evolution of conversational agents. The main idea of this innovative technology is to match user text inputs to a topic, and each topic is associated with a specific rule to generate an output. It began to shift attention to semantic analysis and understanding of sentences. The main limitation of using rules in conversational agents is that they are domain-dependent, which makes them inflexible because they rely on manually written rules for specific domains. With recent advances in machine learning and NLP methods, combined with the availability of computing power, new algorithms have been created to implement "advanced" conversational agents without relying on rules and word matching techniques, which has encouraged the commercial use of conversational agents. In particular, the application of conversational agents has expanded with the emergence of deep learning algorithms. One of the new applications is the development of general-purpose voice assistants (such as Amazon's Alexa, Apple's Siri, Google's Google Assistant, Microsoft's Cortana and IBM's Watson). Voice assistants are typically built into smartphones, smartwatches, dedicated home speakers and monitors, and even cars. For example, when the user says a "wake up" word or phrase, the device activates and the voice assistant starts listening. Thanks to speech understanding, the assistant can then understand the commands and respond to the user's requests, usually by providing pieces of information (e.g., "Alexa,

---

[39] J. Weizenbaum, "ELIZA - a computer program for the study of natural language communication between man and machine", *Communications of the ACM.* Volume 9, Issue 1, 1966, pp. 36-45.
[40] K. M. Colby, S. Weber, F. D. Hilf. "Artificial paranoia," *Artificial Intelligence*, 2(1), 1971, pp.1-25.

what's the weather like today in Paris?", answer: "In Paris, the weather is sunny and it's 75°F"), or by completing tasks (e.g., "Ok Google, play my morning playlist on Spotify"). Nevertheless, speech understanding has proven to be a very difficult automatic task due to the tonal, regional, local and even personal variations of human speech.

Source: D. Jurafsky, J.H. Martin, "Speech and Language Processing," Prentice Hall, 2021.

Conversational agents interact with users through what is called natural language, i.e. spoken and written human language. Natural Language Processing (NLP) is the ability of a computer program to understand human language and is an area of research in Artificial Intelligence (AI). A conversational agent is therefore an AI-powered application that uses speech recognition, automatic natural language processing and text-to-speech to inform, entertain or provide services to users of smartphones and connected speakers. Technically, a conversational agent is associated with several AI components that address the following issues: correctly interpret the user's request, find the appropriate result and transform this result into a response (written or oral depending on the type of conversational agent). When the answer is oral, as in the case of voice assistants, the formulation of this answer is also an essential issue, both in terms of the quality of the information that is provided as a result (i.e., its relevance to the user's request) and the quality of the answer (i.e., the quality of the formulation of the answer by a synthetic voice).

The stages of user interaction. The interaction between the user and the voice assistant goes through several stages[41]:

1. The user "wakes up" the voice assistant with a keyword or activation word (also called "hotword" or "wake word") (for example, "Hey Siri" for Apple Siri or "Ok Google" for Google Assistant). This part of the interaction usually takes place locally on the smart device. Some models of voice assistants can also recognize the user by voice. On some other devices, such as smartphones, before saying the activation word, the user must touch the screen or press a button.

2. The voice assistant collects the voice command which consists of an instruction that the user gives it to perform a task such as turning on the lights, playing music or sending a message. It is also possible for the user to ask a real question to receive specific information such as the weather forecast. This stage relies on Natural Language Recognition algorithms that allow the voice assistant to recognize the language in which the user is speaking while following its grammar.

---

[41] For a presentation of these stages, see also O. Gouliaéva, E. Dosquet, Y. Moysan, *La révolution des assistants vocaux - Comprendre les enjeux et réussir ses stratégies marketing,* Dunod. 2020, p. 11 as well as European Commission, *Final Report - Sector inquiry into consumer Internet of Things*, 2022, points 33-37.

3. The voice assistant usually processes the command/question received by the user in two stages. The first stage is to understand the request and this is normally done by a cloud-based technology. This stage is composed of two sub-stages: (1) voice recognition (Speech-to-Text), which processes the user's voice and translates this audio signal into a natural language sentence, and (2) Natural Language Understanding, which understands the natural language sentence and the user's verbally expressed intentions. The device has the ability to keep a history of transcribed requests, a history of audio requests, and other metadata associated with the request such as the time the request was made. The second stage is to identify the available responses. The voice assistant identifies different options to respond to the user's request. More precisely, once the user's intention has been recognized with a sufficient degree of confidence, the program extracts from the sentence the contextual elements necessary for the execution of the query, relying on information extraction algorithms, third-party services (databases, Internet searches, applications, control of devices connected to the system) and extracted contextual elements (history of the conversation and all the metadata accumulated over time and the user's uses of the tool). The available answers are ranked by the voice assistant based on various parameters: the relevance and availability of the answer, the expected satisfaction of the user, the appropriateness of the answer to the type of device involved, and the alignment with the user's current activity.

4. The voice assistant responds by going through two stages: (1) the synthesis in the form of sentences of the answer given by the machine to the user's request with Natural Language Generation algorithms, and (2) the pronunciation of these sentences automatically generated by the voice assistant through a synthetic voice, with "Text-to-Speech" algorithms that allow the machine to transform a text into sounds by producing its spoken equivalent that can be understood by a human being Specifically, it responds in one of the following ways: (i) letting the user know that it does not understand the request; (ii) triggering an action (e.g., playing a song) or retrieving information identified as appropriate (e.g., the weather); or (iii) responding that it has identified the most appropriate option or providing a list of options from which the user can choose the one that best meets the request. Once the command is executed or the answer is given, if the user does not take any other action, the voice assistant finally goes into standby mode.

It is important to note that the machine learning algorithms employed in each of these user interaction phases learn from each interaction to personalize the user's experience with the voice assistant. When deployed inside a device (smartphone, speaker, vehicle), the voice assistant is on standby and is constantly listening. However, until a specific wake-up phrase is detected, no audio message is transmitted

from the voice receiving device and no operation other than the detection of the wake-up phrase is performed.

Human-computer interaction. Interaction between humans and computer systems is evolving towards natural language-based interfaces. Natural language processing technologies allow users to interact with the system (e.g., operating system, search engine, question-answer system) in their own words, rather than using one of many predefined ways of interaction. In this framework, a text or voice conversation is usually initiated by the user asking a question in natural language and the system answers the question in natural language. There are a number of systems that employ natural language-based interfaces at different levels of complexity. Subsequently, a definition of these systems as well as their respective functionalities and differences is provided.

Interaction with the user: technical challenges. Understanding and generating a dialogue with the user is a complicated task. A dialogue is defined as a sequence of statements, each of which is a unique contribution of a speaker to the dialogue. A statement can consist of one sentence, but it can be as short as one word or as long as several sentences. The structure of the statements has important implications for spoken dialogue. A system must know when to stop talking. For example, if the client interrupts, the system needs to know to stop talking (and that the user could make a correction). A system also needs to know when it should start talking. For example, most of the time, in a conversation, speakers start their statement almost immediately after the other speaker has finished, without a long pause, because people are able (most of the time) to detect when their interlocutor is about to finish speaking. Voice assistants also need to detect if a user has finished speaking, so they can process the statement and respond. This task, called *endpoint detection*, can be very difficult because of the noise and the fact that people often pause in the middle of their turn to speak. A key idea of conversation, originally due to the philosopher Wittgenstein[42], is that each statement in a dialogue is a kind of action performed by the speaker. These actions are commonly called *speech acts* or *dialogue acts*. The 4 main classes[43] are:
- Constatives: the speaker's commitment to something being done (answer, affirm, confirm, deny, disapprove, declare);
- Directives: the speaker's attempts to get the recipient to do something (advise, request, forbid, invite, order, ask);
- Commissives: commit the speaker to a future action (promise, plan, pledge, bet, oppose);
- Acknowledgements: expressing the speaker's attitude towards the interlocutor regarding a specific action (apologizing, greeting, thanking, accepting recognition).

---

[42] L. Wittgenstein, "Philosophical Investigations". (Translated by Anscombe, G.E.M.), *Blackwell*, 1953.
[43] K. Bach, R.M. Harnish, "Linguistic communication and speech acts", *MIT Press*, 1959.

These 4 speech acts are used by general-purpose voice assistants and chatbots. A user who asks a voice assistant to do something (e.g., "Turn it up") issues a *directive*. Asking a question that requires an answer is also a way of issuing a directive: when the system says "*What day in May do you want to travel?*", it is as if the system is commanding the user to answer. On the other hand, a user who states a constraint (such as "*I must travel in May*") issues a *constative*. The speech act expresses an important component of the speaker's intention to say what he or she has said.

Sometimes the conversation is entirely controlled by one participant. For example, a journalist interviewing a chef can ask questions, and the chef answers them. In this case, the journalist has the initiative in the conversation[44]. However, in a normal human-human dialogue, it is more common for the initiative to pass from one participant to another, sometimes answering questions, sometimes asking them, sometimes steering the conversation in new directions, sometimes taking the initiative. This type of interaction is referred to as a mixed initiative. The mixed initiative, while the norm for human-to-human conversations, is very difficult to achieve for voice assistants and chatbots. It's much easier to design voice assistants and chatbots to be passive responders. In question-answer systems, or in search engines, the initiative lies entirely with the user. In these user-initiated systems, the user specifies a request and the system responds. The user can then specify another request. Also, these systems risk constraining the user to whom the chatbot or voice assistant asks a question without giving him the possibility to do anything before answering it. These systems can be very frustrating for the user.

In summary, the interaction of the voice assistant with the user goes through two main stages, namely processing and understanding the user's query and identifying and formulating the response. Therefore, we focus on the issues raised by these two stages: the user request first (1) and the generation of the response second (2).

1. The user query (Speech to Text + NLP)

The user's request is made in two technically distinct phases, resulting in a set of vocal and textual processes that materialize in a "Speech-to-Text" algorithmic system and a "Text-to-Speech" system. Speech-to-Text converts speech to text in natural language, while Text-to-Speech converts text to synthesized speech. It is possible to use pre-established human-like neural voices or create a custom neural voice. In other words, the first stage of the interaction of the voice assistant with the user involves activating the terminal and listening to the request (a) and the second involves understanding the request (b). These different stages use voices, whether human or synthetic, captured by provisions that may fall under the law of personal data, the law of literary and artistic property, or more broadly, personality rights.

---

[44] M. Walker, S. Whittaker, "Mixed initiative in dialogue: An investigation into discourse segmentation", *ACL*, 1990.
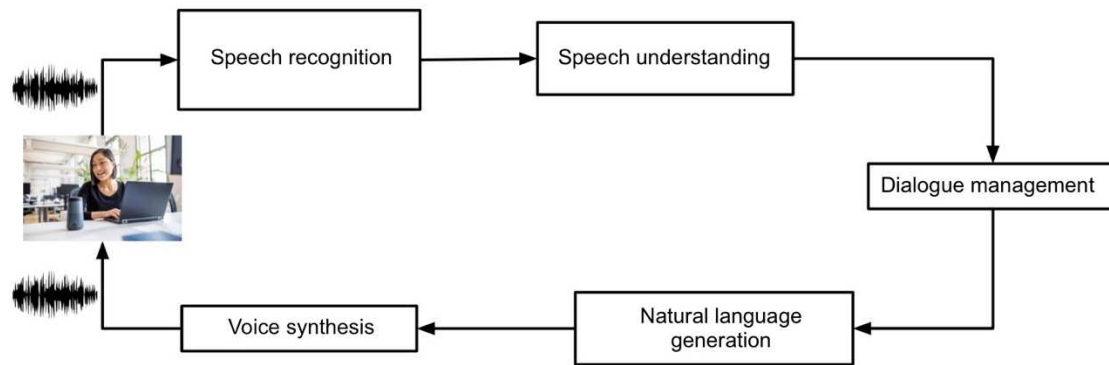
a. Activation and listening

Activation and listening phase. The voice assistant is dormant until it is activated by the user's voice, which will pronounce the word that triggers the activation, or by a routine that is either pre-installed or programmed by the user and which can also be triggered by voice. This voice triggering enhances the user's interaction and experience with the voice assistant, even if they are not fully aware of the effects of the activation. The voice assistant is in the central part of the home (including the vehicle) and takes part in the life of the home as soon as it is activated by intruding into the sphere of intimacy[45].

Upon activation, contextual data (discussions between household members, background noise) are captured by the voice assistant. This capture is not neutral and we immediately perceive the issue of the information and consent of the members of the household, and that of their possible confinement, in this constrained space. The consent of the user and of the members of the household is widely questioned in this context, and at the time of the triggering of the routines. They can perform a succession of predefined daily tasks, sometimes automated, triggered directly by the user's voice or in deferred time once they have been programmed. The voice assistant will then react to a predefined routine by saying the "wake word" (ex: "Hello"). From that moment on, the voice assistant can, inside the home, deactivate the silent mode, adjust the lights, the plugs, the thermostats, give the weather forecast, information on the journey to work, the agenda, broadcast content (music, news, radio, podcast, audio book), and adjust the volume. These routines are already configured on the voice assistant, which does not prevent the user from configuring other, more personalized ones[46]. The various interactions feed the user's profile by putting into

---

[45] "*The voice conveys many characteristics of the individual outside of the speech (the words themselves): emotions, intentions, physical condition, etc. By relying on perception mechanisms, our interlocutors are able to interpret these signals and decipher these states (...) The voice signal thus allows the extraction of numerous and diverse information*": CNIL, *À votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*, coll. White Paper, No. 1, September 2020, p. 4.
[46] HADOPI / CSA, Study supra, p.17.

perspective his life habits, his cultural tastes and his purchases, thus contributing to his confinement, in addition to the use of this information for advertising targeting purposes, which is covered by the newly adopted European texts (see below), or for malicious purposes if the user's installation presents vulnerabilities.

Listening and personal data. The user's voice and those captured in the environment qualify as personal data within the meaning of Article 4 of the GDPR when the person can be identified either by the content of the message itself or if it can be used to deduce, from the voice, the identity, age, gender, socio-cultural origin, state of health or even emotional state[47]. The voice signal emitted by the person is an element of the social interaction that categorizes it according to their qualification. Voice can furthermore be qualified as biometric data within the meaning of Article 4 paragraph 14 of the GDPR[48]. Consequently, the processing of the user's voice during the activation and listening phases is subject to the provisions of the GDPR and the French Data Protection Act (Loi informatique et libertés) as soon as they are applicable. On this point, we refer to the description presented by the CNIL in its white paper "Listening to you", regarding compliance with the requirements resulting from these texts[49]: 1) definition of the personal data processed, 2) definition of the processing and its purposes, its controller and legal basis, 3) choice of the data collected in compliance with the principles of accuracy, proportionality and minimization and their duration of storage, 4) information of individuals, exercise and guarantee of their rights, 5) choice of appropriate measures to ensure the confidentiality, integrity and availability of personal data in the context of concise, clear and transparent information, depending on the sensitivity of the data and the risks identified. Various analyses and recommendations have also been formulated by the European Data Protection Board in its guidelines on virtual voice assistants in order to assist stakeholders in their compliance[50]. In the context of this report, some points to be aware of need to be specified. The first is the difficulty of informing the user; the second is the extent of the listening possible.

Voice interface and user information. The main specificity of the voice assistant consists of its interface with the user through voice commands, without a screen most of the time, which can be an obstacle at first sight to the user's information in the sense of article 5 of the GDPR in the absence of a dashboard, or dedicated applications. It should be recalled that Article 5(1)(a) as well as Articles 12 and 13, clarified by recital 58 of the GDPR, oblige data controllers to inform users of the processing of their personal data in a concise, transparent, intelligible and easily accessible form. So user information is the core of the problem insofar as the configuration and the nature of the interface of the voice assistant do not necessarily allow a good level of information

---

[47] CNIL, White Paper supra, p. 5.
[48] GDPR, art. 4.14: biometric data means "*personal data resulting from specific technical processing, relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data*" - in this sense, EDPB, Guidelines on Virtual Voice Assistants, 02/2021, version 2.0, 7 July 2021, point 31.
[49] CNIL, White Paper supra, p. 47-48.
[50] EDPB, *Guidelines on virtual voice assistants*, 02/2021, supra.

to be provided, which can call into question the legality of the data processing. Especially since voice assistant providers, and more broadly data controllers, face several other obstacles in complying with the transparency requirements of the GDPR[51] that stem from the complexity of the data flows generated and processed. The diversity of the nature of the data processed constitutes a first difficulty, whereas, as the EDPB points out, "*the information provided to the user should match the exact collection and processing that is carried out*". In this regard, it notes that " *while some meta-information is contained in a voice sample (e.g. stress level of the speaker), it is not automatically clear, whether such analysis is performed. It is crucial that controllers are transparent on what specific aspects of the raw data they process* "[52]. It should also be mentioned that the functionalities of the voice assistant can exceed the expectations of users, and this is only the first generation of this type of interface. Voice assistants are more intrusive in capturing contextual and emotional categories of data, when interacting with the user. In addition to this first difficulty, there is also the difficulty of the multitude of actors in the voice assistant ecosystem, which means that particular attention must be paid to devices that allow the addition of third-party functionalities (applications or skills). While some general information may be provided to the user when adding this functionality based on their own choices, the boundaries between the different data controllers involved may be opaquer in normal use of the device. As the EDPB points out, "*the user may not be sufficiently informed how and by whom their data is processed (and the extent to which this processing is occurring) for a specific request*"[53].

The EDPB advocates providing the user with appropriate information in order to guarantee transparency through a better dialogue between human and machine. As for the formal procedures for providing the information, the guidelines recall that, according to Article 12(1) of the GDPR, oral communication of the necessary information is only possible if the data subject requests it, but cannot be used as a default method. In addition, when using an audio message to inform data subjects, data controllers must provide the necessary information in a concise and clear manner with a replay option. If the voice interface remains relevant to inform users of the use of their data, especially for users with disabilities, the EDPB specifies that the information to the user will also have to be provided through an application or a messaging service[54]. It adds that the timing of user notification is also critical (prior to data processing and via a personal interface, such as the smartphone). In the event that the voice assistant provider includes third-party applications in the default configuration - which is mostly the case with general-purpose voice assistants - the EDPB notes that the provider must ensure that users also get the necessary information about third-party data processing. Especially when the general-purpose or specialized voice assistant manages data from a set of stakeholders as it embeds

---

[51] On this point, see EDPB, *Article 25 Guidelines on Data Protection by Design and by Default*, 4/2019, version 2.0 20 October 2020.
[52] EDPB, *Guidelines on virtual voice assistants* supra, point 62.
[53] EDPB, *Guidelines on virtual voice assistants* supra, point 64.
[54] EDPB, *Guidelines* supra, point 57.

services and products (e.g., telecommunications, e-commerce, information technology or web activities), the user must have clear information to choose whether or not the data will be used to build a profile. Beyond that, it is possible to consider that the recognition of a right to parametrisation - the establishment of which will be considered in the third part of this study - could constitute a solution to the processing of data by ricochet.

What about continuous listening? It may seem legitimate for the developer of a voice assistant to want to access information about the interactions that users have with their voice assistant, especially for the purpose of evaluation and continuous improvement of the product. Indeed, it is only in real (or production) conditions that certain behaviours of the system can be observed, as a developer cannot expose his voice assistant to all the situations that may occur during the development phase. However, access to such data must be rigorously controlled, including when it is necessary to perform listening in order to improve the human-machine interaction system. In particular, the activation of the voice assistant must be done in a clear and unambiguous way. The EDPB thus emphasizes the need for users to be able to determine at all times whether the voice assistant " *currently listening on its closed-loop circuit and especially whether it is streaming information to its back-end* "[55] and the need to go through a confirmation of the reception of the voice command (visible light, icons, dashboard, etc.)[56]. In this regard, we recall that the summer of 2019 was peppered with revelations about the practices of major digital actors having these interactions listened to by employees and service providers, and how to access the recordings. The Hamburg Data Protection Commissioner issued an injunction to Google to stop all its activities of having human operators listen to users' interactions with a voice assistant[57]. Subsequently, all the major companies that have performed similar activities (Apple, Amazon, Microsoft and Facebook in particular) have put these activities on hold while they put in place mechanisms for collecting user consent that are satisfactory in terms of the regulations[58]. These methods of informing and expressing consent must be sufficiently explicit and respect the principle of data protection by default enshrined in Article 25, paragraph 2 of the GDPR[59].

Extended listening. While some are deployed on exclusively personal devices, many voice assistants are being deployed in shared environments. In this respect, depending on the places in which the voice assistants physically operate, questions may arise concerning the use of data from uninformed persons, or even the risk of breach of personal or professional secrets. Here again, the terms and conditions

---

[55] EDPB, Guidelines supra, point 63.
[56] EDPB, Guidelines supra, point 51.
[57] https://datenschutz-hamburg.de/assets/pdf/2019-08-01_press-release-Google_Assistant.pdf
[58] V. CNIL, White Paper supra, p. 41.
[59] As an illustration, we can cite, in another case, the conviction of the company Discorde by the CNIL for having set up its application so that it remains active "*even when the user closes the main window, which makes it possible to continue to communicate vocally while not occupying any more space on the desktop of the computer" and whereas "Only a small indicator makes it possible to understand that the application is active*". The CNIL then held that the absence of sufficiently clear and visible information presents significant risks for users, in particular of intrusion into their private lives, and therefore constitutes a violation of Article 25 §2 of the GDPR: CNIL, Deliberation SAN 2022-020, 10 November 2022, Discord INC.

relating to the listening activity must therefore be precisely defined. The EDPB thus underlines the need for the controller to make transparent the type of information that the VVA can obtain about its environment (e.g. listening to other people in the room or to the background music played) as well as to comply with its obligation to inform all users, not only the registered user via his dedicated account, but even those accidentally listened to such as the members of the household present in the home at the time of activation. These requirements will be strengthened if the voice assistant operates on a voice recognition process. Voice recognition aims at uniquely identifying the user on the basis of his/her voice to activate the voice assistant, and at the same time implies subjecting other people (e.g. from the household) to the processing of their biometric data to compare their voice pattern with that of the pre-registered user who wishes to be recognized. In short, this involves, during the activation phase of the voice assistant, filtering the voices (e.g. of people in the household) to identify one of them. If the processing of biometric data is subject to the requirements of Article 9 of the GDPR, then the explicit consent of any person whose data is to be processed, an impact assessment and greater transparency regarding the effects of the collection will be required.

And tomorrow, chatbots: sensitive data, emotional data and risks of manipulation. Automatic speech transcription[60] may reveal sensitive data that in principle cannot be processed unless the consent of the user, or more precisely the data subject, is acquired in accordance with the requirements of Article 9 of the GDPR; this cannot be inferred from mere interaction with the user[61]. However, the question concerns whether the voice itself is sensitive data, which represents a significant issue when considering the new generation of chatbot whose functionalities are extended. There is no reason to object to this approach, if we consider the question from the point of view of biometric data that allow for the unique identification of the person when his or her physical, physiological or behavioural characteristics definitively characterize him or her. Biometric data is also sensitive data because it can be captured remotely without the person's knowledge. For the CNIL, this element of identification of the "speaker" must be subject to regulation[62]. Experiments have been authorized on voice recognition devices by banks in the context of interactive servers[63].

Beyond that, the voice can be used to infer information about the emotional state of the person, and certain technologies are now patented in order to adapt the services offered accordingly[64]. This is the case of Spotify, with its own voice assistant. The

---

[60] Note that voice data extracts are not shared between the different actors, but the content itself may characterize sensitive data. From then on, use is made of the textual transcription of the words, sentences pronounced and movements captured, in order to interpret the meaning of the message transmitted.
[61] On this point, see C. Koumpli, *Les données personnelles sensibles. Contribution à l'évolution du droit fondamental à la protection des données personnelles. Étude comparée: Union européenne, Allemagne, France, Grèce, Royaume-Uni*, Pedone, 2022.
[62] CNIL, Biometrics, https://www.cnil.fr/fr/biometrie
[63] CNIL, White Paper supra, p.9.
[64] EDPB, Guidelines supra, point 48.

streaming service has been granted a patent[65] on new ways to make a voice assistant more responsive to human emotions. The patent describes a voice assistant for a "multimedia playback device" that could, for example, recognize when a user seems sad and acknowledge it sympathetically ("you seem a little down") or offer encouragement ("cheer up"). If a user seems to be in a hurry, the voice assistant could try to keep up by speaking faster or responding with fewer words. Other voice assistants, such as Alexa, are also moving toward empathy[66]. Emotion recognition has a wide range of applications for this voice assistant: it can make conversational AI systems more engaging and provide implicit client feedback that could help voice agents like Alexa learn from their mistakes. In addition, Alexa's simulated emotions give voice app creators the ability to make the voice assistant sound happy and excited or sad and empathetic.

This leads us to wonder about the nature of such data - which we will call "emotional data" - and the regulation of their processing when they could prove to be particularly intrusive. Emotional data allows the individual to be singled out and makes up a body of data specific to his or her physiological and psychological identity, which at least allows it to be qualified as personal data within the meaning of Article 4 of the GDPR[67]. As for considering that this is sensitive data, a reading of Article 9 only allows it to be considered in certain cases, in particular when the emotional data is qualified as health data, relates to sexual life, or falls into the category of biometric data. Apart from these cases, emotion signals could only be subject to the general law on personal data. It could therefore be accepted that emotional data can be processed so that a voice assistant plays content adapted to the emotions felt by its user, provided that such processing respects the principles of the GDPR, and in particular the principle of proportionality. According to these principles, one could think that everything depends on the content proposed in relation to the request addressed. In other words, if the user makes an open-ended request, such as "play a jazz piece", the voice assistant could choose the piece according to the listening context of the user.

Nevertheless, it is possible to consider that allowing such applications could leave the field open to the authorization of much other emotional data processing having a more intrusive impact with regard to the autonomy of the person or even their psychological integrity. The exploitation of human emotions could then reveal the most intimate elements and expose the vulnerability of each person. In other words, a usage-based approach does not seem appropriate given the nature of emotional data. This is particularly true of the prospect of deploying various virtual universes - or other metaverses - aimed at capturing our emotions even more finely through devices such

[65] United States Patent. Systems and Methods for Enhancing Responsiveness to Utterances Having Detectable Emotions. Applicant: Spotify AB. Feb. 18, 2020, https://patentimages.storage.googleapis.com/2a/9d/2d/926b58a2bd956f/US10566010.pdf?fbclid=IwAR2_D4Db8t1AIjtRTW6jSkXcb1ZTFgBvfWWOEtCSgaQBd9wcqNUyKS5WiDw

[66] Adde, V. Rozgnic, "Using Adersational training to recognize speaker's emotions", 21 May 2019: https://www.amazon.science/blog/using-adversarial-training-to-recognize-speakers-emotions

[67] G29, *Opinion on the concept of personal data*, 4/2007, 20 June 2007, WP136, pp. 7-8.

as virtual reality helmets and adapting the virtual environment in which the human avatar would interact accordingly and in real time[68].

Consequently, the minimal protection of emotional data by the law of personal data may seem insufficient in view of the underlying issues, and in particular the risks of manipulation that could result from such processing, including when aimed at evaluating the profile or categorizing individuals. This explains why some[69] consider it necessary to approach this emotional data from the angle of "*sensitive processing*", by making an in-depth reading of Article 5 paragraph 1 of the GDPR, according to which data "must be: a) processed lawfully, fairly and in a transparent manner in relation to the data subject (lawfulness, fairness and transparency); b) collected for specified, explicit and legitimate purposes", while being "adequate, relevant and limited to what is necessary to the purposes for which they are processed". This could impose "*a limitation of processing in case of disproportionate risks - proportionality control - but also a minimization of the data processed as well as a real legitimacy in the purposes of the processing*"[70], so as to avoid processing that is too invasive. But again, these data processing operations should be subject to Article 35 of the GDPR, which prescribes an impact assessment for processing operations presenting "*a high risk to the rights and freedoms of natural persons*", and more specifically any "*processing, in particular through the use of new technologies, and taking into account the nature, scope, context and purposes of the processing, [which] is likely to result in a high risk to the rights and freedoms of natural persons*", and in particular those consisting of a "*systematic and thorough evaluation of personal aspects relating to natural persons […] on the basis of which decisions are taken which have legal effects on a natural person or which significantly affect him or her in a similar way*"[71]. In order to subject such sensitive processing to impact analysis, it is necessary, according to the criteria established by the G29 in its time, that the algorithmic system carries out an evaluation or a rating based on profiling or prediction, and that it is considered that the processing of emotional data "*increases the possible risk for the rights and freedoms of individuals*"[72]. However, what are the rights and freedoms that can be considered in relation to such processing? While it is clear that this reasoning implies a teleological interpretation of the text, it is nonetheless linked to the protection of the rights and freedoms of individuals. Article 3 of the Charter of Fundamental Rights of the European Union, according to which "*everyone has the right to respect for his or her physical and mental integrity*", could be adduced. Nevertheless, the risk of manipulation of emotions in interactions with voice assistants could justify the establishment of new fundamental rights, usefully coupled with the new prohibitions

---

[68] On this point, see in particular A. Basdevant, C. François and R. Ronfard, *Mission exploratoire sur les métavers*, October 2022, p. 85 and p.90 and following - Adde, R. Chatellier, "Métavers : réalités virtuelles ou collectes augmentées ?", *LINC* November 2021.
[69] In this sense, J. Rochfeld and C. Zolynski, "La valeur des émotions : quel régime pour le capitalisme mental?", in *Entre art et technique : les dynamiques du droit, Mélanges en l'honneur de Pierre Sirinelli*, Dalloz 2022, p. 749, spec. p. 759 and following.
[70] J. Rochfeld and C. Zolynski, art. cited supra p. 763.
[71] Idem.
[72] G29, *Guidelines on data protection impact assessment (DPIA) and how to determine whether processing is "likely to result in a high risk" for the purposes of Regulation* (EU) 2016/679, revised Oct. 4, 2017, WP 248 rev. 01, p. 11, criteria 1 and 4.

set by the European texts based on the obligation for honesty, such as the prohibition of targeted advertising, misleading choice interfaces, and transparency with regard to the interlocutor machine. In this respect, it will be necessary to consider the new purposes of human-machine interaction which proceed from the detection of emotions, the construction of reasoning on the basis of this emotional information, but which, in the long run, will be able to generate new emotions specific to an advanced interaction with the machine[73]. Conversational agents can be equipped with emotion detection modules based on linguistic, body and voice modulation cues[74], thus helping to develop appropriate dialogue strategies to match the user's emotions[75]. This amounts to a confinement that can go as far as the cloning of the voice, and the change of emotion of the user when he interacts with the machine.

Such questions are already appearing with the deployment of the first deadbots, these conversational agents likely to figure a conversation with a dead person from the voice or text data left by the deceased person, like Replika[76]. The GDPR does not govern data of deceased persons. Articles 84 to 86 of the amended law No.78-17 of 6 January 1978 on information technology and civil liberties should be consulted to find the principles of the framework. The rights held under this legislation by living persons can be maintained temporarily to organize his or her succession and in particular in two cases. The first, which is identified as "digital death", allows the person to write general or special directives to different people during his or her lifetime. This case remains mostly theoretical. The second is the lack of anticipation of the person who has not decided what should be done with their data. This virtual extension is not without ethical and legal consequences and calls for a legal framework that we assume will be useful for the new data controllers who are the rightful owners. More than ten million people interact with Replika worldwide, and this is just the first of its kind. Having the necessary data to settle the estate of the deceased, the successor has access to personal data and all information useful for the settlement and division of the estate, such as photos. This can feed a deadbot but also an avatar. However, this is a transitional arrangement and responds to the execution of the deceased's instructions. Therefore, it cannot justify the creation of a deadbot or an avatar that would take the place of the deceased over time. A specific right to oblivion, applied "by design", which could be designed in relation to mourning, should probably be circumscribed in order to avoid a virtual extension after death. The stakes are high. Judith Rochfeld specifies to this effect that "*the imperative of oblivion could thus be combined with the social organization of the time of death. The solutions could then no longer depend solely on market forces and the wishes of individuals (or their relatives) but on a social status of mourning when coping with "eternal" digital forces: on a legal consideration of the*

---

[73] CNPEN, *Ethical issues of conversational agents,* opinion No.3, November 2021, p. 20.

[74] L. Devillers, "Le dialogue humain-machine. Intelligence artificielle/intelligence humaine : manipulation et évaluation", *Futuribles* 2019, No. 433, p. 51, spec. p. 52.

[75] R. Chatellier, "Captation des émotions : comment vous le direz pourra être retenu contre vous..." *LINC* April 2018.

[76] For a description of Replika, see Appendix 2 - also, H. Guillaud, "A qui les interfaces s'adaptent-elles?", *InternetActu*, 20 January 2019.

techno-social dimension of death"[77]. Moreover, recourse to the protection of human dignity should not be ruled out when it is a matter of protecting "fragments of the personality" of the deceased[78], provided that proof of their existence can be produced.

b. Understanding the user query

Understanding the user's query and training data. The user's query is interpreted using natural language processing technologies. The intentions of the message are extracted and information variables are identified in order to find the most relevant linguistic pattern to represent the query and thus associate a relevant response to it. One of the issues for this study is to consider how the NLP algorithms designed to understand the user's query were trained and with what data. Corpus-based conversational agents (i.e., text and speech datasets), instead of using hand-constructed rules to generate responses, leverage real conversations between humans to train voice assistants or chatbots. These systems require hundreds of millions or even billions of words to train[79]. Available datasets include transcripts of natural language conversations, such as the Switchboard corpus of American English telephone conversations[80] or the various CALLHOME and CALLFRIEND telephone conversation corpora, available in many languages. Many systems also train on movie dialogues[81], which are similar in many ways to natural conversations. Datasets were also created specifically for training dialogue systems by engaging *crowdworkers* to hold conversations, often having them assume personas or talk about knowledge that was provided to them beforehand. For example, the EMPATHETIC DIALOGUES dataset includes 25,000 crowdsourced conversations based on a specific situation where a speaker was feeling a specific emotion[82]. All of these datasets, while large, do not reach the size of billions of words, so many systems first pre-train on large sets of pseudo-conversations from social networks, such as Twitter[83] and Reddit[84]. Another common technique is to extract possible answers from knowledge sources (such as Wikipedia or news articles) so that a chatbot can tell stories or mention facts acquired in this way.

Finally, once a conversational agent has been put into practice, the turns of phrase that humans use to respond to the chatbot can be used as additional conversational

[77] J. Rochfeld, "Les avatars d'éternité - Vers de nouvelles personnes résiduelles compassionnelles ?", *Mélanges en l'honneur de Catherine Labrusse*, IRJS ed., 2022, p. 754.

[78] Idem.

[79] I.V. Serban, I. V., R. Lowe, P. Henderson, L. Charlin, J. Pineau, "A survey of available corpora for building data-driven dialogue systems: The journal version", *Dialogue & Discourse*, 9(1), pp. 1-49, 2018.

[80] J. Godfrey, E. Holliman, J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development", ICASSP, 1992.

[81] C. Danescu-Niculescu-Mizil, L. Lee, "Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogues". *2nd Workshop on Cognitive Modelling and Computational Linguistics*, 2011; and P. Lison, J. Tiedemann, "Opensubtitles2016: Extracting large parallel corpora from movie and TV subtitles," *LREC*, 2016.

[82] H. Rashkin, E. M. Smith, M. Li, Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," *ACL*, 2019.

[83] A. Ritter, C. Cherry, B. Dolan. "Unsupervised modelling of twitter conversations". *NAACL HLT*, 2010.

[84] S. Roller, E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, E. M. Smith, Y.-L. Boureau, J. Weston, "Recipes for building an open-domain chatbot," *EACL*, 2021.

data for training or finer tuning. It is important to have trust metrics to ensure that these conversational examples are from conversations that are going well and not deceptive or harmful conversations that the conversational agent should not learn, such as hateful conversations. It is also crucial in these cases to remove personally identifiable information[85].

Conversational agents can also be built with hybrid architectures, combining rule-based algorithms and neural networks. This is the case, for example, with systems competing in the Alexa Prize challenge, in which university teams build social chatbots to converse with volunteers on the Amazon Alexa platform, and are scored based on the length and user ratings of their conversations.

Technical issues. Virtual assistant algorithms, used in various forms in systems like Apple Siri, Amazon Alexa, and Google Assistant are designed around a fixed framework. Their objective is to identify a linguistic pattern in the query made by the user in order to know which framework it will be necessary refer to, to seek a relevant answer. Specifically, the goal of the system is to fill in the boxes in the frame (or pattern) with the filler items desired by the user, and then perform the corresponding action for the user (i.e., answer a question, listen to a song, or book a flight). To do this, the system asks the user questions (using pre-specified question templates associated with each slot in each frame), filling in any slot specified by the user. Examples of frameworks are the following:

Slot	Type	Framework of the question
CITY OF ORIGIN	city	"What city are you departing from?"
DESTINATION	city	"Where are you going?"
DAY OF ARRIVAL	date	"On what day do you want to arrive?"

If a user's answer fills several slots, as in the following case: "*I want a flight from Paris to Lyon, one way, departing after 2pm on Tuesday*", the system fills in all relevant slots, then continues to ask questions to fill in the remaining slots, skipping the questions associated with the filled slots. For example, a rule associated with the DESTINATION location of the airplane booking frame, once the user specifies the destination, can automatically enter that city as the default location for the corresponding hotel booking frame. Or, if the user specifies the DESTINATION DAY for a short trip, the system could automatically enter the ARRIVAL DAY. Many areas require multiple frameworks. Different types of inputs trigger different productions,

---

[85] CNIL, White Paper supra, p. 54.

each of which can flexibly fill different frames. Production rules can then change control based on factors such as user input and simple dialogue history such as the last question asked by the system. When the system has enough information, it performs the necessary action (such as querying a flight database) and returns the result to the user.

The objective of these algorithms is to extract three elements from the user's query. The first task is domain classification: is this user talking about airlines, setting an alarm clock or does he want to listen to a song? This classification task is unnecessary for single-domain systems that focus, for example, only on listening to music, but chatbots or general-purpose voice assistants are highly affected by this task. The second is the user's intent: what general task or goal is the user trying to accomplish? For example, the task could be to find a movie, display a flight, or listen to a song that cheers him up. Finally, it is necessary to fill in the slots: to extract the particular frames that the user wants the system to understand from his statement in relation to his intention.

Text and Data Mining. Text and Data mining (TDM) techniques are widely used by private and public entities to analyse large amounts of data (including copyrighted content such as text, images, videos, etc.) in various domains, especially for the development of new applications or technologies. The use of technical TDM on copyrighted works may be subject to different conditions, depending on the legal framework. In generic terms, an act of reproduction is required before TDM can be applied to web-accessible content. As for the benefit of the exception enshrined in Directive 2019/790 Copyright in the Digital Single Market[86] transposed into the Intellectual Property Code by the Ordinance of 21 November 2021[87], it is subject to the *opt out* of the right holder when outside the context of research. Article 4.3 of the directive, taken up by Article L. 122-5-3-I of the IPC, states that this *opt out* must be express and must be expressed "*in an appropriate manner, in particular by machine-readable processes for content made available to the public online*"[88]. Article L. 122-28 of the IPC, introduced by the Decree of 23 June 2022[89], specifies the terms and stipulates that "*The opposition mentioned in III of Article L. 122-5-3 does not have to be reasoned and may be expressed by any means. In the case of content made available to the public online, this opposition may, in particular, be expressed by means of machine-readable processes, including metadata, and by the use of general terms of use of a website or service*".

---

[86]Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.
[87] Ordinance No. 2021-1518 of 24 November 2021 supplementing the transposition of Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the digital single market and amending Directives 96/9/EC and 2001/29/EC.
[88] V. A. Bensamoun et Y. Bouquerel, *Transposition des exceptions de fouille de textes et de données. Enjeux et propositions,* Mission report presented to the CSPLA, 15 December 2020.
[89] Decree No. 2022-928 of 23 June 2022 amending the Intellectual Property Code and completing the transposition of Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the digital single market and amending Directives 96/9/EC and 2001/29/EC

In this context, a machine-readable solution that streamlines the communication of available TDM rights and licenses for copyrighted content online is needed to facilitate the development of these applications and reduce the risks of legal uncertainty for the actors implementing these techniques. This solution also aims to optimize the ability of TDM actors to lawfully access and process valuable content on a large scale. The W3C (Word Wide Web Consortium) Community Group "TDM Reservation Protocol"[90] has recently proposed a new protocol called TDMRep. The purpose of this protocol is, among other things, to allow a rights holder to declare its choice regarding text and data mining (TDM) on Web resources under its control, allowing recipients of this declaration to adjust their *scraping* behaviour or to enter into a separate agreement with the rights holder that satisfies all parties. Specifically, this protocol provides three complementary techniques for expressing the choices of rights holders. These correspond to different situations and technical skills that an editor may have: (1) the TDM file on the origin server is a mechanism for declaring site-wide rights holder choices in a file hosted on the origin server of the Web content that a TDM agent wishes to exploit; (2) the TDM header field is a mechanism for declaring a choice in an HTTP response; (3) the TDM metadata in HTML content is a mechanism for declaring a choice embedded in HTML content.

2. Answer generation (NLP + Text-to-Speech)

The generation of the answer by the voice assistant goes through three main stages. First, the system must find the most relevant answer to the user's query (a). Next, the system must generate the answer in natural language, using natural language generation algorithms (b). Finally, the system must express the generated sentence in a synthetic voice (c). The main technical and legal issues related to these three stages should be highlighted.

a. Answer generation (NLP)

The natural language generation module of the voice assistant produces the statements that the system passes on to the user. Most systems are frame-based, that is, they tend to use template generation, in which all or most of the words in the sentence to be spoken to the user are pre-specified by the virtual assistant designer. The sentences created by these templates are often called *prompts*. The models can be completely fixed (e.g., "Hello, how can I help you?") or may include some variables filled in by the generator depending on the user's request (e.g., "What is the weather like in Paris today?", "The weather in CITY_NAME is WEATHER" -- "The weather in Paris is cloudy"). A more sophisticated generation component can condition the exact

---

[90] TDM Reservation Protocol (TDMRep), Final Community Group Report - https://www.w3.org/2022/tdmrep/

context to produce turns of phrase that seem much more natural. Current voice assistants implement a combination of both approaches.

Single answer. Most corpus-based chatbots and voice assistants generate their responses to a user's turn to speak in context, either by *information retrieval* methods (using information retrieval to extract a response from a corpus appropriate to the context of the dialogue) or by *natural language generation* methods (using a language model or neural network-based algorithm to generate the response based on the context of the dialogue). In both cases, the systems basically generate a single appropriate response to the entire conversation so far. This is why they are often called question and answer systems. Corpus-based chatbot or voice assistant algorithms are thus inspired by the algorithms of question-answer systems, which also focus on single answers while ignoring longer-term conversational goals. As highlighted in the CNIL White Paper on voice assistants[3], "*providing only a single answer raises questions about the choice of result sources*".

Search engine and answer engine. The problem of the single answer lies in the algorithms implemented by search engines and question-answer systems.

Regarding the operation of search engines, the initial neural search, assumed much simpler methods, such as exact matching of strings. String matching was used to filter a certain term on a website or in a document. Exact string matching returns all documents that contain the search term, in no particular order. To increase the relevance of search results, it was possible to use keyword matching. According to keyword matching methods such as *tf-idf*, the terms of a document are weighted according to their relevance in the corpus. These methods are fast and language independent. In contrast, keyword matching works on the simplifying "*bags of words*" principle. This means that it does not preserve the word order in a document. Word order is important for the meaning of a sentence because it contributes to our understanding of the syntactic and semantic relationships between words. In modern search engines, semantic (or "neural") search algorithms are implemented. As such, instead of comparing a query (for example, a string or an image) directly to the database, it goes through a neural network that has been pre-trained on millions of data. The neural model has learned to encode a query as a high-dimensional vector. This high dimensionality allows neural models to better capture the meaning or semantic value of a query. The latest neural language models are all based on the transformer architecture.

Transformers

A "transformer model" is a neural network architecture that can perform complex tasks such as automatic text or image generation with reduced training costs. In particular, it consists of

---

transformer layers capable of modelling sequential dependencies between different elements (images, pixels, words) to reduce the training time of these models. For example, transformer models are used to solve all sorts of tasks in automatic language processing. The pipeline of a transformer model connects a model with its necessary preprocessing and post-processing stages. By default, this pipeline selects a particular pre-trained model that has been refined for a specific NLP task. Three main stages are required when passing text to a transformer model: (1) the text is pre-processed into a format that the model can understand; (2) the pre-processed inputs are fed to the model; (3) the model predictions are post-processed, so that they can be given meaning.

The transformer architecture was introduced in June 2017. Initial research focused on machine translation tasks. This was followed by the introduction of several influential models, including:

- *June 2018*: GPT, the first pre-trained transformer model, used for fine-tuning on various NLP tasks, which achieved state-of-the-art results.
- *October 2018*: BERT, another large pre-trained model, this one designed to produce better sentence summaries.
- *February 2019*: GPT-2, an improved (and larger) version of GPT that was not immediately made public for ethical reasons.
- *October 2019*: DistilBERT, BART and T5, large preformed models using the same architecture as the original transformer model.
- *May 2020*: GPT-3, an even larger version of GPT-2, capable of good performance on a variety of tasks without the need for fine tuning.

All of the transformer models mentioned above (GPT, BERT, BART, T5, etc.) have been trained as language models. This means that they have been trained on large amounts of plain text in a self-supervised manner. Self-supervised learning is a type of training in which the objective is automatically calculated from the model inputs. This means that humans are not needed to label the training data. This type of model develops a statistical understanding of the language on which it has been trained. An example task is to predict the next word in a sentence after reading the previous n words. However, one must be careful with these models that work by observing the statistical relationships between words and sentences, but do not understand their meaning.[91]

Source: D. Jurafsky, J. H. Martin, "Speech and Language Processing," Prentice Hall, 2021.

The answer engines interpret the question, evaluate the available options and provide the best possible answer. Specifically, a question answering system accepts a question and then returns not a document, but an answer. The term "extractive question answering" refers to the system highlighting one or more passages (from one or more documents) as answers. An answer engine normally provides a single answer, which is the most relevant of the possible answers identified by the algorithm.

Evaluation. Evaluation is crucial in the design of chatbots and voice assistants. These are usually evaluated differently because they have different goals: generalist voice assistants have to accomplish a task like booking a flight or playing a song; chatbots

---

[91] https://www.nature.com/articles/d41586-021-00530-0

have a different kind of goal, like being pleasant to users. Chatbots are evaluated by humans, who give them a score. This can be the human who talked to the chatbot (participant evaluation) or a third party who reads the transcript of a human/chatbot conversation (observer evaluation). In the participant evaluation, the human evaluator chats with the model for a few minutes and scores the chatbot on 8 dimensions capturing conversational quality: avoiding repetition, interest, meaning, fluency, listening, curiosity, humanity and engagement. Observer ratings use annotators to review the text of a complete conversation; for example, raters may rate the consistency of each statement.

b. The voice expression of the answer: Text-to-Speech

There are two main issues in expressing the answer: first, once the system has found the most relevant answer to provide to the user, it formulates that answer in natural language, using natural language generation algorithms. This text is then translated into an audio file using text-to-speech algorithms. The second challenge concerns the generation of the synthetic voice, which uses different speech synthesis algorithms. Therefore, both challenges must be considered.

Text-to-Speech. Text-to-Speech (TTS) aims to synthesize intelligible and natural speech from text. It has long been a subject of research in AI, natural language processing and speech processing. The development of a TTS system requires knowledge of languages and human speech production, and involves multiple disciplines, including linguistics, acoustics, digital signal processing, and machine learning. A TTS system consists of a base of three elements: a text analysis module, an acoustic model and a vocoder. The text analysis module converts a text sequence into linguistic features, the acoustic models generate acoustic features from the linguistic features, and then the vocoders synthesize the waveform from the acoustic features.

Several techniques. Different techniques are proposed for the generation of synthetic voice (WaveNet[92], Tacotron[93], DeepVoice[94], Google Duplex[95]). The most common method of producing a synthetic voice is called concatenative synthesis. This method searches for phonemes (i.e., distinct sound units in a specific language) and combines these "pieces of speech" recorded from human voices to produce synthesized speech. In general, this approach does not produce good voice quality because of the way it assembles these units. The system draws from an audio database, segmenting waveforms that create enough variation in speech that the human ear does not

---

[92] https://www.deepmind.com/blog/wavenet-a-generative-model-for-raw-audio
[93] https://google.github.io/tacotron/
[94] S. O. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman, S. Sengupta, M. Shoeybi, "Deep Voice: Real-time Neural Text-to-Speech," *Proceedings of the 34th International Conference on Machine Learning*, ICML, p.195-204, 2017.
[95] https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

perceive them realistically. Important advances in this field have been proposed with neural networks. The use of a neural network allows us to take an ordered set of phonemes and transform it into a set of spectrograms.[96] The neural network uses these and selects the appropriate spectrograms with the frequency bands that more accurately articulate the acoustic features that the human brain uses to systematically understand and organize speech. A neural vocoder then translates these spectrograms into speech waveforms, which allows the voice to be reproduced in a natural way. These approaches can also model the parameters of a voice such as intonation, rhythm, and timbre, by auto-encoders and generative adversarial networks. There are three main techniques for concatenative synthesis:

(1) Domain-specific synthesis: this involves stringing together pre-recorded words or phrases to form complete statements. This method is commonly used for simple, repetitive use cases such as weather reports or pre-takeoff instructions. However, this solution is limited to the words and phrases for which it has been programmed.

(2) Unit selection synthesis: it draws from a large database of audio clips of pre-recorded speech of an actor previously transcribed and broken down into phonemes, syllables, morphemes, words, expressions and sentences. These units are then indexed and collated to determine the best sequence for the target sentence.

(3) Diphone synthesis: uses a database containing all the diphones present in a specific language.[97] Once drawn, the prosody is superimposed on these units using digital signal processing techniques. The result is not as good as unit selection but better than domain specific synthesis.

As highlighted by the CNIL[98], this topic is important in relation to *deepfakes* that exploit neural networks to manipulate or generate sound content with a high potential for deception.

Technical issues in Text-to-Speech

With the development of deep learning, neural network-based TTS (neural TTS) is proposed, adopting deep neural networks as the basic model for speech synthesis. A key element in building an automatic speech recognition model is to train it on a range of different voices, so that it can learn a variety of acoustic frequency profiles and different ways of expressing phonemes, which are the shortest units of speech. For example, the WaveNet model for directly generating the waveform from linguistic features can be considered the first neural model of TTS. Other models such as DeepVoice rely on statistical parametric synthesis, but enhance it with neural network-based models. Some end-to-end models (e.g., Tacotron, Deep Voice and FastSpeech) are proposed to simplify text analysis modules and directly support input characters and phonemes, and to simplify acoustic features with mel spectrograms.

---

[96] A spectrogram is a visual rendering of the spectrum of frequency bands of a signal.
[97] For example, Spanish has about 800 diphones while German has 2500.
[98] CNIL white paper supra, p.7.

Specifically, the TTS system extracts prosody embeddings from a reference audio file and uses them as additional input to the neural network. During learning, a reference audio file is used, and during inference, another reference audio file helps to synthesize speech with a similar prosody. Style tokens can increase the ability and variation of TTS models to learn different types of styles. Each token can learn different representations of prosody, such as different speech rhythms and emotions. The model can use a reference audio file to assist and extract prosody representations, or simply choose one or more style tokens to synthesize speech. Adaptive TTS models have the objective of synthesizing a user's voice. These models are the basis for voice adaptation, voice cloning and personalized voice systems. In this framework, a source TTS model (usually trained on a multi-speaker speech dataset) is usually adapted with some adaptation data for each target voice.

These models are at the origin of the synthesized voices of voice assistants and conversational agents. For example, Amazon Alexa synthesized each utterance in a dataset 32 times, randomly sampling 32 voice profiles from 500 collected from volunteers in the lab. Some of the data may come from Cleo, an Alexa skill that allows multilingual clients to contribute to the training of new language models by responding to voice prompts with open-form statements. This data is not made public. In the case of Apple Siri, for iOS 11, a female voice talent was chosen with the aim of improving the personality and expressiveness of Siri's voice. Hundreds of candidates were evaluated before one was chosen. Subsequently, more than 20 hours of speech were recorded to build a new synthesized voice using a TTS model based on deep learning. Sonos' voice assistant, *Sonos Voice Control*, leverages the voice profile of "Breaking Bad" actor Giancarlo Esposito. Esposito spent approximatively 40 hours in the studio recording thousands of phrases and sentences that were then used as training data for the voice model. He also read material specific to the Sonos voice assistant, even though these phrases are not used in 1:1. In addition, Sonos is using voice recordings from its beta user community to form Sonos Voice Control.

Manufacturing the synthetic voice and protecting the human voice. It is therefore appropriate to question the extent of protection that should be accorded to recordings of human voices when employed in synthetic voice contexts. Notably, the synthetic voice comprises a compilation of a series of words, phonemes, and diphones, pronounced with varying intonations by one or more persons whose voices will serve as models. The service provided by a person to make these recordings may be covered by an employment contract and thus be subject to remuneration. In this respect, it is customary for the voice of an actor, used to constitute the synthetic voice of voice assistants, to be covered within a contractual and remunerated framework, which is generally calculated by the line or by the hour, as seen in cases of dubbing or recordings of digital audio books. Beyond that, the question arises as to whether

this recording can also give rise to the artist's performing right or whether it can be covered by the protection conferred by personality rights.

Protection under the artist's performing right. A first avenue to explore is to consider the protection of the person who made these sound recordings with regard to the artist's performing right within the meaning of Article L. 212-1 of the IPC. Several conditions must therefore be fulfilled. The first condition results from the letter of the text according to which "*the performer is the person who represents, sings, recites, declaims, plays or performs in any other manner a literary or artistic work, a variety act, a circus act or a puppet act*". It must be deduced that "*the performer is the person who represents [...] in any other way a creative or intellectual work*"[99] and that "*there is no interpretation without a work*"[100]. This first condition can therefore pose a problem when it comes to the recordings made for the constitution of the synthetic voice of a voice assistant. This assumes that they result from the interpretation of a creative or intellectual work, rather than mere words, snippets of conversation or readings of texts that lack the creative aspect outlined in Article L. 212-1 of the IPC. Cas law concerning a person's image rights[101] could be relevant in contexts where an individual solely "lends" his voice to the voice assistant. In addition, the nature of the performance of the person whose voice is being recorded must be considered. The qualification of artist supposes in fact that an interpretation is carried out which translates a personal expression of the interpreter in his or her performance, independently of the merit of the interpretation and the fame of the artist, which excludes from the protection any person whose role remains purely technical[102]. Therefore, the recording of the reading of a work made by a "teller" (performer, journalist) can be qualified as interpretation within the meaning of Article L. 212-1 of the IPC as soon as it is not linear, and that it reveals the personality of the performer. It should also be noted that dubbing or voice-over artists who lend their voice to a main character are qualified as performers[103]. The recording of such a reading must be governed by an artist's contract in order to provide for the conditions of the transfer of rights and the related remuneration. According to Article L. 212-3 III of the IPC, it must be appropriate and proportional to the actual or potential economic value of the rights assigned, taking into account the performer's contribution to the work as a whole and all other circumstances of the case, such as market practices or the actual exploitation of the performance, in the case of monetizable exploitation.

It is necessary to clarify what happens if the recording of this interpretation is compiled with other sound extracts in order to create the synthetic voice of a voice assistant. The first question therefore concerns the application of exclusive rights. In this respect, the InfoPaq judgment accepted that parts of a work can be *"protected by copyright as*

---

[99] A. Lucas, A. Lucas-Schloetter and C. Bernault, *Traité de la propriété littéraire et artistique*, Lexis Nexis, 5th ed. 2017, No.1335.
[100] A.E. Kahn, "Droits voisins du droit d'auteur. Droit des artistes-interprètes ", *Jurisclasseur Propriété littéraire et artistique*, 2022, No.1.
[101] See in particular Paris, 18th ch. sect. D, 12 February 2008, No.C07/02721.
[102] A.E. Kahn, supra, No.43 and following.
[103] Collective agreement on the rights of dubbing performers, signed on 6 January 2005 and extended on 12 March 2005 - see also Paris, ch. 2, Nov. 20, 2015.

long as they contribute as such to the originality of the whole work", after recalling that "*Article 2(a) of Directive 2001/29 provides that authors have the exclusive right to authorise or prohibit the reproduction of their works in whole or in part*"[104]. The Pelhman decision confirms this analysis by stating that the exclusive right conferred by Article 2, c) of Directive 2001/29 of 22 May 2001 on the phonogram producer to authorize or prohibit the reproduction of his phonogram "*allows him to object to the use by a third party of a sound sample, even a very brief one, of his phonogram for the purpose of including that sample in another phonogram*". However, the Court of Justice clarified that this provision must be interpreted in the light of the Charter of Fundamental Rights of the European Union, from which it follows that the reproduction right is conditional on the fact that the sample is not included in another phonogram "in a modified form unrecognisable to the ear"[105]. In the event that the sound sample used by the creator of the new work is recognizable upon listening, the Court accepts that its reproduction may fall within the exception of quotation enshrined in Article 5, 3, d. of Directive 2001/29, read again in the light of the Charter, "*provided that use has the intention of entering into dialogue with the work from which the sample was taken*"[106]. The wording adopted by the Court of Justice may lead to the acceptance that such limitations to this exclusive right presuppose that the extract is exploited to constitute a new work insofar as they are justified by the freedom of the arts - freedom of creation - protected under Article 13 of the Charter[107]. Therefore, these limits would hardly apply to extracts of sound recordings compiled in order to constitute the synthesized voice of a voice assistant, all the more so when it is a process with an exclusively commercial purpose. There remains the second question concerning the basis for the remuneration due for the exploitation of excerpts of recordings of performances by several performers, which could be based on an annual flat-rate remuneration, on the model of the practices concerning sound banks offered for purchase.

Protection as a complementary artist. In the absence of exploitation protected under performing rights, one can usefully consider the qualification of "complementary artist", defined in opposition to the performer, and for which two criteria are traditionally used: the absence of personality in the performance, and the interchangeable or

---

[104] CJEU, 16 July 2009, Infopaq International A/S v Danske Dagblades ForeningInfopaq, case C-5/08, points 32 and 38. It should be noted that, while the Court of Justice accepts that "*words as such do not constitute elements covered by the protection*" (point 46), it holds that, "*given the requirement of a broad interpretation of the scope of the protection conferred by Article 2 of Directive 2001/29, the possibility may not be ruled out that certain isolated sentences, or even certain parts of sentences in the text in question, may be suitable for conveying to the reader the originality of a publication such as a newspaper article, by communicating to that reader an element which is, in itself, the expression of the intellectual creation of the author of that article. Such sentences or parts of sentences are, therefore, liable to come within the scope of the protection provided for in Article 2(a) of that directive"* (point 47).
[105] CJEU, 29 July 2019, Pelham GmbH, Moses Pelham, Martin Haas v. Ralf Hütter, Florian Schneider-Esleben, case C-476/17, point 39.
[106] CJEU, Pelham, supra, point 72.
[107] On the citation exception, however, see CJEU 1 December 2011, Eva-Paria Painer, case C-145/10, point 136, which held that "*the issue of whether the quotation is made as part of a work protected by copyright or, on the other hand, as part of a subject-matter not protected by copyright, is irrelevant*". However, it should be noted that some authors have pointed out that the scope of this decision should not be too broadly understood, as the solution can be explained by the facts of the case (the reproduction of a photograph in a press article which constitutes a "citing structure" even if it is not qualified as a creative or intellectual work protectable by copyright): M. Vivant and J.-M. Bruguière, *Droit d'auteur et droits voisins*, Dalloz, 2018, 4th ed, No.657; also A. Lucas, A. Lucas-Schlotter, C. Bernault, *Traité de propriété littéraire et artistique*, Lexis Nexis, 5th ed. 2017, No.465.

substitutable character of the performance. A decision relating to a voice recording made by an actor as part of an employment contract for an advertising campaign illustrates the reasoning used by the case law in this area[108]. Noting the lack of interpretation of a creative or intellectual work, the judges rejected the qualification of performer since the personality of the artist was not expressed in his performance insofar as the voice was not distinguishable from those of other actors who had lent their voices to the character in previous recordings. It should also be noted that the brevity of the performance and the professional nature of the performance are irrelevant in this respect[109].

Perspectives. Various projects are currently underway, notably for dubbing artists, to organize a voluntary collective management of the exploitation of these voice recordings. The creation of a database would make it possible to identify the "teller" holder of the right to remuneration and consequently facilitate management companies in organizing this collective management using the tools available at their disposal. Given the progress of algorithmic voice processing techniques, it seems necessary to support such projects and to invite the involvement of organizations defending the interests of the profession in order to plan, supervise and remunerate the use or borrowing of artists' voices used as a matrix for these systems.

Protection of the voice as a personality right. If the synthetic voice is based on characteristics derived from natural persons, it is appropriate to consider their protection under personality rights insofar as the case law accepts that the voice constitutes an attribute of the personality, "*a kind of sound image*"[110]. As previously noted[111], the human voice is an element of privacy protected under Article 9 of the Civil Code. This protection arises because the voice makes it possible to identify the person, through the content of the message it carries, but also as such when one can deduce from the person's voice his or her identity, age, sex, socio-cultural origin, state of health and even emotional state. The voice can also be seen as "*an attribute of fame that can be monetized like an image*"[112]. Some decisions go so far as to accept an autonomous right to the voice[113], which some see as a full-fledged personality right[114]. Following a movement in favour of designation of the voice as an asset similar to that concerning the exploitation of the image, the person is thus recognized as having a right to exploit his or her voice, which is likely to be commercialized, as

---

[108] Versailles, Oct. 9, 2008 and the observations of J.-M. Bruguière, *Propriétés intellectuelles* 2009, No. 31, p. 173.
[109] See not. Paris 31 May 1996 and Cass. Civ. 1st, 6 July 1999 for an appearance of 2 seconds out of the 15 seconds of the commercial.
[110] TGI Paris, 19 May 1982, *D.* 1983, jurisprudence p. 183 concerning the unauthorized broadcast of Maria Callas' voice - Paris, Jan. 22, 2001, *D.* 2002, p. 2375, note A. Lepage.
[111] E.g. TGI Paris, 3rd ch. 1st sect. 27 Sept. 2004, *CCE* 2004, comm. 153, obs. C. Caron - see in the United States, concerning the commercial use of the voice in disregard of a right of publicity, *Midler v. Ford Motor Co.*, 849 F. 2d 460 (9th Cir., 1988): imitation in a series of commercials of Bette Midler's voice, seeking to make it appear that she was singing herself and thus profiting from her fame; *Waits v. Frito-Lay*, 978 F. 2d 1093 (9th Cir., 1992), cert. denied, 113 S. Ct. 1047 (1993); imitation of Tom Waits' voice also for commercial purposes (commercial sound-alike).
[112] J.-M. Bruguière and B. Gleize, *Droits de la personnalité*, Ellipses, 2015, No.183 - Also, C.-A. Maetz, *La notoriété. Essai sur l'appropriation d'une valeur économique*, PUAM, 2010, p. 251-252.
[113] TGI Paris, Sept. 27, 2004.
[114] See in this sense Versailles, 9 October 2008, supra. - Adde, D. Huet-Weiller, "La voix et le contrat : le contrat sur la voix ", *RTD civ.* 1982, p. 497.

attested to by practices in the advertising sector or with regard to professional athletes and trainers[115]. However, a person may challenge the use of the information for commercial or advertising purposes if he or she did not intend to participate in such activities or claim commercial harm in the form of lost earnings[116]. This means that an individual can object to the use of his or her sound identity to generate a synthetic voice, or even to having his or her voice transformed or extended by means of algorithmic processing in the absence of his or her authorization.

Synthetic voices: towards new models? We are now seeing the development of a new market for synthetic voice. Several companies offer this type of service, such as Voxygen or Acapella. The uses are diverse, and allow for the production of films, documentaries or television programs with deceased actors or historical figures, such as the documentary "Judge Pétain", made from press articles and official minutes of the trial held in 1945, using a sound modelling process developed by IRCAM that allowed for the reconstitution of the voices and intonations of the main protagonists As Nicolas Obin, a researcher at IRCAM's Science and Technology of Music and Sound Laboratory, points out, "*We are capable of artificially reproducing a human voice, like the voice of actor André Dussolier, which can now be produced by a machine and create phrases that the actor never uttered, or that of Marilyn Monroe*". He specifies that "*Ircam works on the characteristics of the voice and is thus able to hybridize voices, for example by mixing a human voice with that of a lion. With advanced software, it is also possible to modify the attributes of a voice: from a neutral recording, it is possible to add emotions to a voice, to change it from anger to joy or sadness. The challenge of this research is to create more expressive voices, the next stage in speech synthesis…*"[117]. The work carried out by these teams also focuses on the adaptation of the synthesized voice to its context of use. It is then a question of adapting the synthetic voice in particular not to be, as it is now, rather monotonous during the reading of a work. Moreover, Nicolas Obin notes that although today "*the voice of voice assistants is often designed on a model of neutrality in order to adapt to everyone*", this work allow us to envisage other models to adjust the vocal identity of the voice assistant by using specific intonations or accents. Beyond that, voice variations could promote new forms of human-machine interactions by adapting to the listening context or even to the emotions of the user of the voice assistant, which is now possible thanks to the deployment of chatbots based on Affective Computing techniques.

---

[115] See, for example, Article L. 222-2-10-1 of the Sports Code regarding contracts concluded with a professional athlete for the commercial exploitation of his or her image, name or voice - Adde, M. Serna, "La voix et le contrat : le contrat sur la voix", *CCC* 1999, chron. No.9.
[116] *Traité Droit de la personnalité*, ed. J.-C. Saint Pau, LexisNexis, 2017, No. 1233.
[117] Interview by H. Guillaud, "Hypervoix 1 : concevoir les interfaces vocales", *Internetactu.net*, 24 May 2019.

Affective computing

Affective computing aims at training computers using human emotions or affects. This is a multidisciplinary field of research, in which sociology, psychology, computer science, physiology, mathematics and linguistics make important contributions. *Affective computing* includes the procedure of acquiring physiological and behavioural signals, processing these signals, extracting *features* and classifying or predicting them. These extracted features are used to train a data model, which can then classify or predict outcomes in various domains. This technology uses a variety of techniques to serve three purposes: to detect emotions and "reason" on the basis of this affective information, to generate new affective expressions; to give rise to human-machine interactions based "*on the notion of theory of mind, which designates the cognitive ability of an individual to attribute unobservable mental states (intention, desire, belief, emotion) to oneself or to other individuals*", in other words on the fact that when "*a human being perceives an emotion in another, it is a subjective perception of his or her own*"[118].

Source: R. Arya, J. Singh, A. Kumar, "*An inquiry of multidisciplinary domains contributing to affective computing*", Computer Science Review, Volume 40, 2021.

The deployment of these techniques which can "*give multiple inflections to the synthetic voices, for example by allowing them to suggest, to incite, or even to make the listener feel guilty*"[119] coupled with the development of *Affective Computing* based on the processing of the interface user's emotions to approach the model of human conversation raise questions about the possible manipulation of the user of the voice assistant. This raises questions about the regulation of these tools and uses, both because of their ethical[120] and legal[121] issues. On this last point, it should be noted that, for the time being, Article 52 of the proposed European Regulation establishing harmonized rules on artificial intelligence, dated 21 April 2021 (AI Act), requires providers of AI systems intended to interact with natural persons to ensure that these systems are designed and developed in such a way that natural persons are informed that they are interacting with such systems[122], which amounts to imposing here only an obligation of transparency[123].

---

[118] National Digital Ethics Steering Committee, CNPEN, *Ethical issues of conversational agents*, opinion No. 3, Nov. 2021, p. 20.
[119] N. Obin, Interview supra.
[120] On this point, see the opinion of the CNPEN, supra.
[121] See in particular J. Rochfeld and C. Zolynski, "La valeur des émotions : quel régime pour le capitalisme mental ? ?", in *Entre art et technique : les dynamiques du droit, Mélanges en l'honneur de Pierre Sirinelli*, Dalloz, 2022, p. 749.
[122] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Articificial intelligence (Artificial Intelligence Act), COM/2021/206 final - see also in this sense the *Ethical Guidelines for Trustworthy AI* from the High Level Expert Group appointed by the European Commission, published in 2019, according to which "*AI systems should not present themselves as human beings to users; when interacting with an AI system, human beings have the right to be informed. This aspect implies that AI systems must be identifiable as such*" (point 78).
[123] Some go so far as to advocate a ban on these emotional recognition systems, whose impact on the psychological integrity of individuals is highlighted by the Council of Europe (Council of Europe, *Towards a regulation of AI systems. International perspectives on the development of a legal framework based on Council of Europe standards in the field of human rights, democracy and the rule of law*, DGI 2020(16), p. 26, point 20), subject to certain specific uses for research or health protection purposes (see in particular EDPB-EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)*, 18 June 2021, point 35) or the

Co-creation opportunities involving the user's voice may be in the near future, which promises to be more fertile ground for machine-learning speech synthesis to personalize the human-machine relationship. For the time being, machine learning has difficulty handling personalized voice interfaces because they cannot be based on user data alone. Associating it could then be a decisive step in favouring voice assistant profiles with more singular, more unexpected vocal interfaces[124], likely to erase the existing gap between the human and the machine's level of perception. This is why the creation of a synthetic voice with quantities of recordings of the user's voice is one way to achieve this. The modelling of emotions and human reasoning announces a configuration in which "*Tomorrow, the interface will be you*"[125]. Such prospects raise many questions: will this non-human other, which is constructed within the framework of an anthropomorphic approach involving the "user" and the "used", with the voice of this user, become the mirror of his own verbal behaviour and his own verbal limits, thus suggesting a rapid confinement?[126] Isn't this the evolution of an interaction approach that runs the risk of impoverishing language and creation if the voice assistant is not powerful enough? *"Indeed, the more the similarity between human and machine increases, the more the deviation from the model can generate a feeling of uneasiness or even apprehension"*[127]. Beyond that, it raises questions about the model of society to be promoted and the human will to restrain technical models in order to avoid reaching such extremes in the near future. In this case, many avatars will be created in all creative universes with the voice of the users, given that many start-ups already offer to create the user's own voice avatar. Should we then consider that all users will eventually be interpreters, having the vocation to replace professional interpreters[128] in order to give voice to virtual assistants and to accentuate the idea of a digital cloning of the person? In this same perspective, are we able to accept avatars designed for eternity with the voices of deceased people?[129] The same technology can be used for voice prostheses, in other words voice cloning[130]. These possibilities of co-creation of personified synthetic voices to improve the empathy of the human-machine relationship will certainly have an impact on the protection of the voice by a copyright that should be reinvented in the light of human-machine co-creations, and a performers' right that should be adapted.

---

demonstration that the algorithmic system would strengthen the autonomy of the person (see CNCDH, *Opinion on the impact of artificial intelligence on fundamental rights*, 7 April 2022, A-2022-6, p. 14 and following).

[124] H. Guillaud, supra, p.2.

[125] Expression by L. Julia, *Hypervoix. Promises and illusions of new human-computer interfaces*, FING, 2019, p. 33.

[126] FING, *Hypervoix supra.*

[127] Idem, p.34.

[128] F. Vallet, " Les droits de la voix (2/2) - Quelle parole pour nos systèmes ?", *LINC*, 27 June 2019.

[129] J. Rochfeld, "Les avatars d'éternité - Vers de nouvelles personnes résiduelles compassionnelles ?", *Mélanges en l'honneur de Catherine Labrusse*, ed. IRJS, 2022, p. 754.

[130] See on this point, interview with N. Obin, CNIL White Paper, supra, pp. 33 and following.

c. The content of the vocal response: the issue of the communication of a work in sound form

The answer of the voice assistant can be of a plural nature depending on the question asked by the user and the Text-to-Speech technique used for the operation of the interface, ranging from the simplest and completely contextual answer to a more comprehensive discourse depending on the development of the request and its precision. As such, within the framework of the present report, one of the points in question proceeds from the answer brought by the voice assistant which could consist of the communication of a work in sound form by the technique of vocal synthesis (Text-to-Speech), which amounts to determining if this communication could give rise to the right of reproduction and the right of communication to the public.

Voice communication and reproduction right. With regard to the right of reproduction, it is necessary to consider the conditions set out in Article L. 122-3 of the IPC, according to which "*reproduction consists of the material recording of the work by all processes which can communicate it to the public in an indirect way*". These conditions seem to be fulfilled in this case. Firstly, it is indeed a material recording within the meaning of the text since the recording can be carried out by any process, and that "*the nature of the media used for the recording is of no consequence*"[131], which includes reproduction on a computer memory[132] or a remote storage space[133]. This will be the case for the communication of a work in sound form generated by the technique of voice synthesis (i.e. the artificial production of the human voice) which supposes, for this purpose, the setting in digital format of, for example, the text, stored either on the hardware (e.g. the connected speaker) or in the cloud. Secondly, the purpose of the work is indeed to be communicated to the public, here to the user of the voice assistant. Moreover, the transition from written to oral remains of no consequence[134].

It is then necessary to verify whether the exception of temporary reproduction could be applied. In this respect, the conditions set out in Article L. 122-5, 6° of the IPC provide that "*the temporary reproduction having a transitory or incidental character, when it is an integral and essential part of a technical process and that it has for only object to allow the licit use of the work or its transmission between third parties by the way of a network calling upon an intermediary; however, this temporary reproduction, which can relate only to works other than the software and the databases, must not have to have an economic value of its own*". As for the provisional character, it should be noted that the retention "*cannot exceed what is necessary for the proper functioning*

---

[131] A. Lucas, A. Lucas-Schloetter, C. Bernaud, *Traité de propriété littéraire et artistique*, 5th ed. 2017, No.259.
[132] CJEU, 16 July 2009, Infopaq, Infopaq International A/S v. Danske Dagblades Forening, case C-5/08.
[133] TGI Paris, 14 August 1996 or TGI Paris, 5 May 1997.
[134] Cass. Civ. 1st, 15 October 1985.

of the technical process"[135]. Consequently, the copy must be deleted as soon as it has fulfilled its technical function. It therefore comes down to checking whether it is transitory or incidental. In the case of the reproduction of the work to produce an oral communication by means of synthesis, the incidental character of the copy will only exist if the purpose is to communicate the work by sound and not to reproduce it. The conditions laid down by the Court of Justice in the NLA judgment will then be fulfilled because the reproduction will have neither existence nor autonomous purpose[136]. It should be noted that, for the time being, communication by synthetic voice is still in a preliminary state (the reading of a work currently consists of a pre-recorded podcast-type application using human voices), but it is possible that technical advances will allow, within a reasonable time frame, the reading of a work directly by the voice assistant through Text-to-Speech techniques. However, it is not yet possible to decide how the text or speech generated afterwards will be stored. It can be considered, for example, that the operator will have an interest in storing the recording of the generated synthetic voice. This would enable it to be replayed upon the request of other users, thus avoiding the need for new algorithmic processing. In this case, the exception available for temporary copying would not apply. Finally, it will be necessary to verify the absence of economic value of the copy, which could be debated[137], especially if one considers the resulting data flow.

Voice communication and right of communication to the public of works. With regard to the right of representation, Article L. 122-2 of the IPC stipulates that "*the representation consists of the communication of the work to the public by any process*". This must apply if the vocal synthesis allows an indirect communication of the work to an audience and all the characteristic elements of the work are communicated by the intermediary of this technical process. Furthermore, in accordance with the provisions of Article 3 of Directive 2021/29, such communication consists of the representation of the work in such a way that everyone can access it from a place and at a time individually chosen by them. As a result, the right holder can oppose the communication of the work by this process in the absence of authorization. It should also be noted that the author may oppose this communication if the voice representation of the work infringes the right to respect for its integrity, particularly in the case of representation by synthetic voice of poor quality.

---

[135] CJEU Infopaq, supra.
[136] CJEU 5 June 2004, Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and others, case C-360/13.
[137] In this respect, see CJEU Infopaq, supra, which dismisses the independent economic significance provided, firstly, that the implementation of these acts does not allow for an additional profit, going beyond that derived from the lawful use of the protected work and, secondly, that the acts of temporary reproduction do not lead to a modification of the work.

B - The voice assistant as a new interface

Beyond the algorithmic system, the voice assistant must also be designed as an interface allowing privileged, more natural access to many services offered either by the voice assistant provider or by third parties, which the user can choose from an application store. A description of the universe of the voice assistant (1) allows us to better understand how the regulatory framework is gradually seeing it as a new interface, even as a particular form of intermediation (2).

1. Description of the voice assistant universe

Voice assistant and terminal. A voice assistant is software that runs on a physical device, or terminal equipment. The technology of a voice assistant is represented by a program (software) that is installed on a stand-alone device or terminal (hardware). Specifically, a typical voice assistant has two layers: firstly, the software layer (which includes the frontend, i.e., the "face" of the application, consisting of the user interface and voice recognition technology, and the backend that runs services such as AI algorithms to process a question, understand natural language, locate an answer); secondly, the hardware layer (the endpoint that includes computers, smartphones, smart speakers, other IoT devices, etc.). Voice assistant technology can benefit and continually improve with the latest advances in software and hardware. Many such hardware solutions are closed (e.g., smart speakers); the software may also be proprietary, as are most general-purpose voice assistants (Amazon Alexa, Apple Siri, Google Assistant).

Voice applications. Voice applications are features that enhance the user experience. They can be activated manually or directly from the speaker with the voice command. These applications are designed for all the tasks you want to be able to ask the voice assistant to perform, such as providing news, recipes, playing music, finding the best transportation for a given trip, etc. Each platform has a different name for its voice applications. Amazon uses the term *skills*[138], Google and Siri opt for the term *applications*[139]. It is important to note that each platform has its own built-in applications, such as to ask for the time, the weather and playing a piece of music. Voice applications created by third-party developers can be invoked using a specific phrase or, if the virtual assistant allows it, can be invoked implicitly, without triggering the key phrase. For example, in the invocation "*Hey Google, talk to <application name>*" it is explicitly stated which application is requested. However, the invocation "*Hey Google, what's the weather like today in Paris?*" implies by the context of the request what service the user wants, i.e. an application to predict the weather. In the case of an implicit invocation, the voice assistant will be able to decide which

---

[138] https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html
[139] https://play.google.com/store and https://www.apple.com/fr/app-store/

application among those installed and authorized by the user it will use to answer his request, in case multiple installed applications could accomplish the same task. In case the voice assistant does not have a third-party application installed and indicated as preferred by the user for a certain task, it will preferably rely on its own built-in applications. The combination of accounts linked to a paid offer and the use of voice applications ideally allows for optimal and personalized use of the speakers.

Application stores and their challenges. An application store is a type of digital distribution platform for computer software called applications, often in a mobile context. Application stores typically take the form of an online store, where users can browse the different categories of applications, view information about each application (such as reviews from users who have used the application), and acquire the application (purchasing the application, if necessary, or downloading applications offered for free). The selected application is offered for automatic download. This is followed by its installation. All apps, app updates, app packs, in-app purchases and in-app events submitted to the app store are reviewed by the app store owner. The major voice assistant manufacturers all have their own app stores (e.g., Alexa Skills, Apple App Store, Google Play, Microsoft Apps), where users can select the apps they want to install and associate with their profile. This operation allows the voice assistant to "call" and activate the application among all the applications available in the store. Often this includes the user logging in with their account on the application to allow the voice assistant to use that application at the user's request. When a new app is built, it is necessary for developers to use the guidelines and resources provided by the app store owner to ensure that the app review process goes smoothly before it is published in the store. This could lead to voice assistant operators imposing their rules on third parties, thereby intensifying the complexity of the validation process of third-party applications before they are published in the application stores of these same operators. As highlighted in the United States House of Representatives' investigation into digital markets, market actors are then concerned about their difficulties in gaining access to key functionality needed to build their applications, such as unprocessed user commands[140].

The voice assistant ecosystem. As illustrated in Figure 1 of the House of Representatives' report[141], voice assistants can be viewed as a user interface that enables exchanges between computing devices through a person's voice command. Specifically, these computing devices can take different forms such as smart speakers (Google Home, Amazon Echo Dot) or mobile devices and computers. The voice assistant, connected to these devices through the network, provides an interface to address requests to these devices with a voice command. To these devices, we can also add home automation and IoT devices such as lighting, thermostats, security

---

[140] House of Representatives, *Investigation of the competition in digital markets*. Majority Staff Report and Recommendations. CP 117-8, Part 1, July 2022.
[141] House of Representatives investigation, supra, p. 101, prepared by the Subcommittee based on Hyunji Chung, Jungheum Park & Sangjin Lee, Digital Forensic Approaches for Amazon Alexa Ecosystem, 22 DIGIT. INVESTIGATIONS S15 (2017), https://dfrws.org/wp-content/uploads/2019/06/paperldigitallforensiclapproacheslforlamazonlalecosystem.pdf

monitors and even kitchen appliances that are connected to voice assistants through the cloud, a digital work environment on the Internet to store data and run applications (i.e. applications activated from the application store).
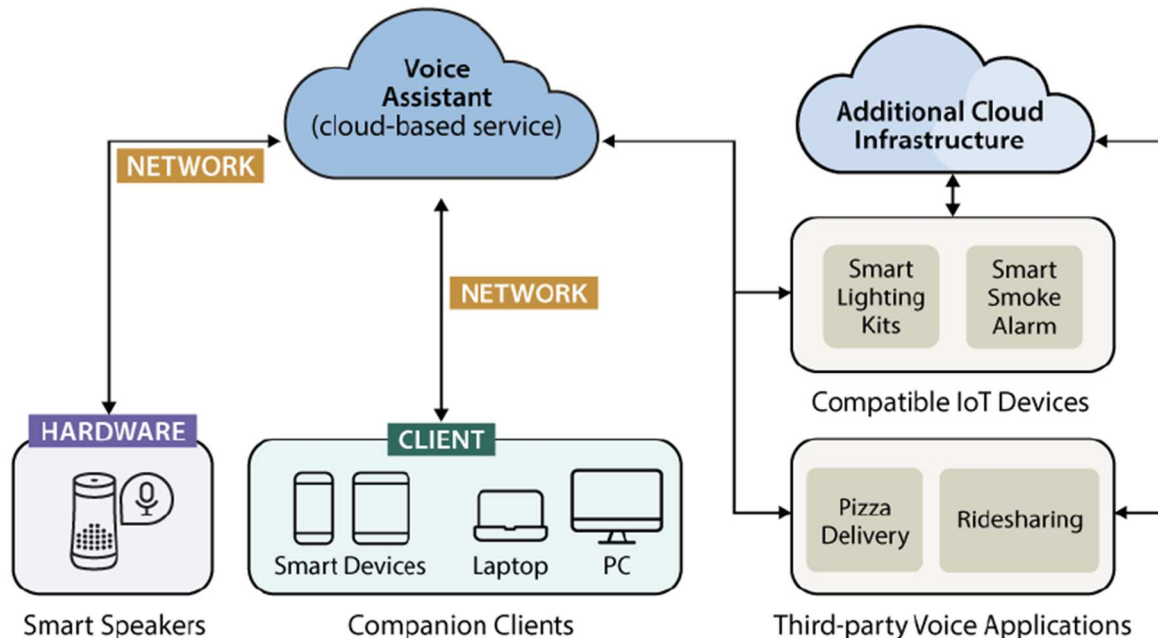
Figure 1 - Source: House of Representatives, *Investigation of competition in Digital Markets*, CP 117-8, Part 1, July 2022, p. 101

Two dimensions stand out in this world of voice assistants: (1) the dimension that takes into account the hardware and software components around the voice assistant, and (2) the dimension that takes into account the applications and application stores connected to the voice assistant.

For dimension (1), the hardware component includes connected speakers, smartphones and home automation devices, while the software component includes the voice assistant itself as software (see above). The voice assistant allows you to send voice requests to the hardware components. As highlighted by the House of Representatives report: *" The smartphone and smart speaker are the two main portals for voice assistants […]. Market participants emphasize that smart speakers represent an essential "hub" or gateway for smart homes and are driving voice-assistant adoption […]. A voice assistant platform vendor can expand its ecosystem by adding IoT devices and voice applications. Both IoT devices and voice applications can be first-party - owned by the voice assistant platform vendor - or third-party, if the vendor has set up services to allow for manufacturers to create voice assistant-enabled devices"*[142].

---

[142] House of Representatives investigation, supra, p. 101-102.

Regarding dimension (2), in this context, the voice assistant represents a privileged interface to access the proprietary services of its provider. This privileged interface allows a preferential access and a promotion "by default" of the applications belonging to the same universe as the voice assistant. For example, the Amazon Alexa voice assistant favours listening to music with the application (*skill*) that is part of its own universe, i.e. Amazon Music. These applications do not require explicit integration by the user, unlike third-party applications that must be explicitly integrated among the applications that the voice assistant can access and launch using voice commands.

Issues. In this environment, several issues can result from the interaction of the voice assistant with its own universe, such as network effects, cross-referencing of user usage data, self-preference and lock-in strategy.

Network effect. Like many online platforms, the voice assistant market benefits from network effects. Specifically, the more users there are on a platform, the more third-party devices and applications are available, which draws more users to the platform. These network effects for voice assistant platforms are amplified by machine learning tools and Artificial Intelligence. Improvements in automatic natural language processing algorithms and AI are expected to improve the quality of voice assistants and contribute to broader adoption.

Use and cross-referencing of data. User usage data includes all data about actions taken by the user in interaction with the voice assistant. This data can obtain an improvement in the performance of the voice assistant, for example an improvement in the understanding of the user's request. More specifically, the collection and cross-referencing of this data strengthens the ecosystem of certain operators, allowing these actors to improve their own services, and at the same time to obtain a better knowledge of the user's profile. This profile becomes more complex and includes not only the data explicitly provided by the user but also those collected during his interactions with the voice assistant. Inferences on these data thanks to the implementation of data cross-referencing algorithms allow the acquisition of a deep knowledge of the user and his preferences. This opens up the issues of personal data protection, already mentioned above. The result is a more attractive offer in terms of answers and services provided to the user and a better "market position". In particular, the interaction with the user becomes more natural because the voice assistant has learned to know the user through his past interactions.[143] In addition, the cost to store this data can be lower when the voice assistant operator can leverage other components of its universe, such as the cloud, allowing the operator to offer lower cost data storage and service offerings. To summarize, the use of usage and user profile data may favour the position of voice assistant operators in the market.

---

[143] In addition, the user improves his ability to interact with the voice assistant.

Self-preference and self-promotion. Voice assistant operators can use all this data to promote other services available in their universe, such as an operator's application in the app store. This triggers self-preferential issues on the two dimensions mentioned below. Thus, the voice assistant operator's speaker is preferentially used to interact with the voice assistant and the applications that will be preferred to respond to the user's requests will be those of the same operator (e.g., the Amazon Echo Dot speaker employs the Amazon Alexa voice assistant to play music with the Amazon Music application). The lack of a screen in the interaction with voice assistants as well as the challenges associated with single response on these systems may lead to favouring the services of these operators by default.

2. The voice assistant seen by the regulations as a new form of intermediation

As described above, the voice or virtual assistant is designed as a new interface at the centre of the relationship between the voice assistant provider and the user of the voice assistant, allowing the user to access related services, offered either by the voice assistant provider or by third party companies. It is also an additional feature for accessing an ecosystem, and can even become the entry point for accessing a set of digital services offered by its provider. Consequently, the voice or virtual assistant characterizes a new form of intermediation for which it is advisable to think of a relevant legal framework. This is the purpose of recent reforms of European Union law regarding digital markets and services. Several texts now explicitly target voice or virtual assistants in the light of this new form of intermediation.

Introduction of voice assistants in EU law. The first text to have expressly referred to voice assistants was the Regulation of 20 June 2019 promoting fairness and transparency for companies using online intermediation services, known as Platform to Business[144]; this clarification was made at the very end of the negotiations of the text in order to capture a form of interface that was beginning to be deployed in Europe, and particularly in Germany. Recital 11 states that the online intermediation services covered by this Regulation include, "*by way of example, e-commerce marketplaces, including collaborative ones on which business users are active, online software applications services, such as application stores, and online social media services, irrespective of the technology used to provide such services. In this sense, online intermediation services could also be provided by means of voice assistant technology*". A few months later, it was the European Union's consumer law that took up the issue on the occasion of the cross-sectoral reform of the acquis carried out by Directive 2019/2161, known as the Omnibus Directive[145], particularly to specify that

---

[144] Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.
[145] Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules.

the form imposed to respect the information obligations prescribed by the text will have to be reviewed when the contract is concluded through "voice operated shopping assistants" when it is a voice command[146]. Its Recital 41 thus states that Article 8(4) of Directive 2011/83/EU concerns remote contracts, such as those entered into by "*the telephone, via voice operated shopping assistants or by SMS*"[147].

New form of intermediation and functioning of digital markets: the DMA. The Digital Markets Act[148] includes virtual assistants in its scope. Article 2.12 of the regulation defines them as "*software that can process requests, tasks or questions, including those based on audio, visual, written input, gestures or motions, and that, based on those demands, tasks or questions, provides access to other services or controls connected physical devices*". The text allows us to cover the multiple facets of voice assistants, namely both as a new user interface and, above all, as a gateway to multiple markets.

It should be remembered that the DMA's approach is intended to be comprehensive and forward-looking, halfway between the law of restrictive competition practices and the regulatory approach[149]. The text applies to the largest gatekeepers who enjoy a hegemonic and stable position that gives them considerable economic power. This reform is based on the fact that these operators connect many user companies and many end users, which they take advantage of to increase their power over the ecosystem of which they are the epicentre, in particular thanks to access to vast quantities of data. This can lead to serious imbalances and be the source of unfair practices and conditions for the companies that use their service as well as for the end users. The text therefore imposes *ex ante* obligations on these operators in order to prevent them from monopolizing the market to which they provide access, by preventing the entry of competitors or by imposing their own operating rules in their own interest (e.g., prohibition of so-called self-preference practices consisting of favouring one's own services, or of combining and cross-referencing data between services). Other obligations aim at opening up the market in order to facilitate access to third parties (e.g. principle of non-discrimination, interoperability, access and portability of data)[150].

With the adoption of the DMA, the virtual assistant is included in the definition of core platform services, whereas this was not the case in the early stages of negotiation of the text. This addition, proposed by the European Parliament's Committee on the

---

[146] On this issue, see K. Sein, "Concluding Consumer Contracts via Smart Assistants: Mission Impossible Under European Consumer Law?", *EuCML* 2018, 179.

[147] See also Communication from the Commission, *Guidance on the interpretation and application of Directive 2011/83/EU of the European Parliament and of the Council on consumer rights* (2021/C 525/01), point 3.2.2.2.

[148] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on questionable and fair market in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Market Regulation).

[149] In this sense, see J.-C. Roda, "Un an de droit de la concurrence dans l'univers numérique", *CCE* 2022, study No.12, No.2 and M.-A. Frison-Roche and J.-C. Roda, *Droit de la concurrence*, Dalloz, 2022, No.665 &s. - also S. Abiteboul and J. Cattan, *Nous sommes les réseaux sociaux*, Odile Jacob, 2022, p. 182 and following.

[150] On these different obligations, see in particular M. Debroux, " ature, objectifs et régime des obligations imposées aux contrôleurs d'accès : Des obligations *ex ante* et *per se*, vraiment ?", *Concurrences* No.3-2022, pp. 66-72. Adde, M. Le Roy, *La loyauté des plateformes à l'égard des consommateurs*, Dalloz bibliothèque des thèses, 2023, No.274 and following.

Internal Market and Consumer Protection and then validated by the Council, can be explained by the conduct, during the negotiations, of the European Commission's sectoral inquiry on the Internet of Things. This investigation highlighted the challenges and significant concerns of actors that the practices of the major voice assistant providers are negatively impacting competition, innovation and consumer choice in the IoT sector, and was further compounded by the announcement of an investigation launched by the European Commission into Google Assistant in light of its default installation on Android-powered devices[151].

If virtual assistants fall within the scope of the DMA as core platform services, it will nevertheless be necessary to consider the qualification of gatekeepers in order to subject these core platform services to the obligations set out in the text. The logic of the DMA is to focus on these large companies insofar as they provide services. Once this qualification is accepted, the second stage will be to know what are the core platform services provided by these companies. For this reason, gatekeepers are considered to be those that meet the following quantitative criteria. These are companies with a very high turnover or market value: 7.5 billion euros or more in annual sales in Europe over the last three years or 75 billion euros or more in market capitalization over the last year. Since 90% of digital companies (about 10,000) are considered SMEs unable to capture the market, restoring access to the market means imposing obligations on the remaining 10% and considering that the 10,000 companies in question are users of the services provided by gatekeepers. In the text, the services falling within the scope are referred to as "core platform services" and are 10 in number.   These include intermediation services (such as marketplaces, application stores); search engines; social networks; video sharing platforms; online messaging; operating systems (including connected TVs); cloud services; advertising services (such as ad networks or exchanges); web browsers and virtual assistants. These 10 services are problematic in the sense of the DMA, and are therefore subject to the obligations set out in the text, if they are provided in at least three European countries; if they register more than 45 million Europeans per month and if they use the services of 10,000 professionals per year for the last three years. In short, once the companies have been designated as gatekeepers, the Commission will have to list all the services that meet these conditions, which will then be called core platform services.

The DMA text posits other definitions for virtual assistants, including clarifying what is meant by the number of unique end users, namely those, who, "*engaged with the virtual assistant in any way at least once in the month, such as for example through activating it, asking a question, accessing a service through a command or controlling a smart home device*". Under the DMA, the number of user companies is targeted at the "*number of unique developers who offered at least one virtual assistant software application or a functionality to make an existing software application accessible*

[151] https://www.competitionpolicyinternational.com/eu-probes-googles-voice-assistant/

through the virtual assistant during the year". Therefore, the question of the application of the DMA to core platform services remains dependent on compliance with the thresholds by the various virtual assistants present on the market. However, the inclusion of voice assistants during the negotiations on the text makes it possible, at the very least, to perceive the potential of these actors in the making, and the possibility of covering them with the help of the qualitative criteria laid down by the text in Article 3.8, as it is currently difficult to know whether they meet the quantitative thresholds in the absence of numerical data. If the company in question does not meet the thresholds referred to in Article 3(2), the Commission may, pursuant to Article 3(8) of the DMA, designate as gatekeeper any company that (1) has significant market power in the internal market; (2) provides a core platform service that constitutes a major access point enabling user companies to reach their end-users; and (3) enjoys a strong and sustainable position in its business, or is likely to do so in the near future. For this purpose, "*the Commission shall take into account some or all of the following elements, insofar as they are relevant for the undertaking providing core platform services under consideration:*

(a) the size, including turnover and market capitalisation, operations and position of that undertaking;

(b) the number of business users using the core platform service to reach end users and the number of end users;

(c) network effects and data driven advantages, in particular in relation to that undertaking's access to, and collection of, personal data and non-personal data or analytics capabilities;

(d) any scale and scope effects from which the undertaking benefits, including with regard to data, and, where relevant, to its activities outside the Union;

(e) business user or end user lock-in, including switching costs and behavioural bias reducing the ability of business users and end users to switch or multi-home;

(f) a conglomerate corporate structure or vertical integration of that undertaking, for instance enabling that undertaking to cross subsidise, to combine data from different sources or to leverage its position; or

(g) other structural business or service characteristics".

Thus, it is clear from Article 3.8 of the DMA that the ecosystemic approach should guide the Commission in qualifying gatekeepers, which is particularly important for capturing the activity of digital services covered by the text[152]. Furthermore, it should be noted that, if the provision of the voice assistant is not considered a core platform service, the company qualified as a gatekeeper may have its activities regulated by the text in accordance with this same systemic approach. In fact, for certain prescribed obligations, *"the DMA does not reason by service but by ecosystem of the company concerned".* For example, according to Article 5.2, *"the gatekeeper will not be able to cross-use the data acquired in the framework of its different services"*[153].

---

[152] On the need to regulate companies and their ecosystems, see J. Toledano, *Gafa. Reprenons le pouvoir !,*, Odile Jacob 2020, p. 128 and following.
[153] D. Bosco, "A propos du concept de contrôleur d'accès", *Concurrence*, No.3-2022, p. 61, spec. p. 63.

Beyond that, it can be observed that the logic of the text allows for the identification of obligations that would be applicable in general to voice assistants if the company providing the voice assistant is designated as the gatekeeper. However, given the diversity of voice assistant business models, it seems necessary to recommend specific obligations for some of them, depending on their position on the market and their manoeuvrers to circumvent the rules. The drafting of delegated acts provided for in the DMA to enable the Commission to extend the obligations laid down in the text according to the mechanism envisaged in Article 8 cannot create this asymmetry, since they are designed as implementing acts for the regulations, which cannot go beyond the letter of the text. If the *ex ante* regulation approach is supported by the text, it remains symmetrical. This explains why some people are calling for an evolution of the future text towards a logic closer to regulatory law, allowing these obligations to be adapted "customized" to each operator in a more efficient and responsive manner, so that Union law does not remain only halfway effective[154].

New form of intermediation and data sharing. The proposed Data Act Regulation[155], which is currently being negotiated, also targets voice assistants by adopting, almost identically, the definition of the DMA ("*software that can process demands, tasks or questions including based on audio, written input, gestures or motions, and based on those demands, tasks or questions provides access their own and third party services or control their own and third party devices*"). As part of the European data strategy alongside the Data Governance Regulation, which defined a governance framework for industrial data trading, this text should help open up a market for the data produced by connected devices. Its horizontal legislation model will, like the DMA, affect different heterogeneous business models, products and services. In this context, the Data Act highlights the essential role of voice assistants in the Internet of Things, and the value of data sharing for the benefit of the consumer or between companies, within whose scope voice assistants will fall (Article 7). The goal: to foster the development of a *" competitive offer of aftermarket services, as well as broader data-based innovation and the development of products or services unrelated to those initially purchased or subscribed to "*.

Proposed Data Act Regulation, Recital 22

" Virtual assistants play an increasing role in digitising consumer environments and serve as an easy-to-use interface to play content, obtain information, or activate physical objects connected to the Internet of Things. Virtual assistants can act as a single gateway in, for example, a smart home environment and record significant amounts of relevant data on how

---

[154] K. Favro and C. Zolynski, "DSA-DMA : l'Europe au milieu du gué ", *Dalloz IP IT* 2021, pp. 217-223.
[155] Proposal for a Regulation of the European Parliament and of the Council of on harmonised rules on fair access to and use of data (Data Act), COM(2022) 68 final, 23 February 2022.

users interact with products connected to the Internet of Things, including those manufactured by other parties and can replace the use of manufacturer-provided interfaces such as touchscreens or smart phone apps. The user may wish to make available such data with third party manufacturers and enable novel smart home services. Such virtual assistants should be covered by the data access right provided for in this Regulation also regarding data recorded before the virtual assistant's activation by the wake word and data generated when a user interacts with a product via a virtual assistant provided by an entity other than the manufacturer of the product. However, only the data stemming from the interaction between the user and product through the virtual assistant falls within the scope of this Regulation. Data produced by the virtual assistant unrelated to the use of a product is not the object of this Regulation".

The voice assistant, a new form of intermediation according to the meaning of the DSA? The Digital Services Act[156] aims to regulate the practices of digital services in order to correct malfunctions by making digital service operators accountable, similar to financial regulation. The objective of the text is to protect fundamental rights, to ensure a secure environment through the establishment of a new governance, and to respond to issues that evolve on a daily basis. To this end, the European Union has chosen to implement cross-sectoral and coherent regulation of digital services. The text is applicable to a number of actors whose business models and underlying issues are nonetheless highly heterogeneous. The DSA thus promotes regulation through transparency of intermediation services. The horizontal approach of the DMA is only imperfectly reflected in the DSA, which is content with a double vertical approach aimed at online platforms and search engines.

It should also be noted that the DSA does not cover the same function with respect to intermediation. It applies to all online intermediaries who offer their services (goods, content or services) on the domestic market and, more specifically, to the platform's intermediation activity *per se* (e.g., its activity of moderating third-party content or recommending third-party content through the use of a recommendation algorithm). It is intermediation of another nature that the DMA covers, that of acting as gateway to a market of the operator, an activity that allows it to define the rules of organization of that market. The voice assistant does not seem to qualify as an intermediary within the meaning of the DSA, in respect of which it remains a simple interface, which explains why it is not formally covered by the text.

Nevertheless, the spirit of this text opens the field of possibilities with regard to the intention and the obligations formulated therein. With regard to intent, Recital 12 states that " *In order to achieve the objective of ensuring a safe, predictable and trustworthy online environment, for the purpose of this Regulation the concept of 'illegal content' should broadly reflect the existing rules in the offline environment. In particular, the concept of 'illegal content' should be defined broadly to cover information relating to*

---

[156] Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act).

illegal content, products, services and activities. In particular, that concept should be understood to refer to information, irrespective of its form, that under the applicable law is either itself illegal, such as illegal hate speech or terrorist content and unlawful discriminatory content, or that the applicable rules render illegal in view of the fact that it relates to illegal activities. Illustrative examples include the sharing of images depicting child sexual abuse, the unlawful non-consensual sharing of private images, online stalking, the sale of non-compliant or counterfeit products, the sale of products or the provision of services in infringement of consumer protection law, the non-authorised use of copyright protected material, the illegal offer of accommodation services or the illegal sale of live animals. In contrast, an eyewitness video of a potential crime should not be considered to constitute illegal content, merely because it depicts an illegal act, where recording or disseminating such a video to the public is not illegal under national or Union law. In this regard, it is immaterial whether the illegality of the information or activity results from Union law or from national law that is in compliance with Union law and what the precise nature or subject matter is of the law in question". The operation of the voice assistant as an intermediary between third-party user companies and users could be qualified as practices governed by the text aimed at, for example, parametrisation (Article 25) and other misleading interfaces or misleading advertising. Nevertheless, if these practices identified in the operation of voice assistants are not covered by the qualification of online platform services, namely by the DSA, they will have to be dealt with by the market.

A particular issue: conversational agent and AI regulation (AI Act). Insofar as conversational agent are based on an algorithmic system of automatic learning, they could be covered by the legal framework under construction concerning the regulation of artificial intelligence. In this respect, the proposed Regulation laying down harmonized rules on artificial intelligence, known as the AI Regulation[157], does not, *per se*, qualify conversational agents either as very high-risk systems - which are prohibited from being deployed in the internal market - or as high-risk systems - which are governed by the compliance rules set out in the text. In its press release dated 21 April 2022, the European Commission unambiguously states the classification of conversational agents: they are artificial intelligence applications or systems that present only a limited risk to the fundamental rights of individuals[158]. The text then only imposes transparency obligations on the AI system provider. For example, Article 52 provides that *"providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use"*. It is therefore a matter of protecting the end user from a possible confusion of status during his interaction with the machine[159]. Note that the Boltering Online Transparency Act of California Senate Bill 1001 (known

---

[157]Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts (COM/2021/206 final).
[158] European Commission, *A Europe fit for the digital age. Commission Proposes New Rules and Actions for Excellence and Confidence in Artificial Intelligence,* Press Release, 21 April 2021.
[159] On this confusion of status, see CNPEN, Opinion No.3, *Ethical issues of conversational agents*, November 2021, pp. 7-8.

as the "Bot Act")[160] also provides that "*it is unlawful for any person to use a bot to communicate or interact with another person in California online, with the intent to deceive the other person about his or her artificial identity for the purpose of knowingly misleading the person about the content of the communication in order to induce the purchase or sale of goods or services in a commercial transaction or to influence a vote in an election*". As for Article 52(2) of the AI Regulation, it is aimed more specifically at emotional recognition systems, specifying in this regard that "*users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences*". These provisions, which are still under negotiation, show that it is the particular nature of the human-machine interface and the manipulation issues that may result with regard to the end user that justify the establishment of a - simple - obligation of transparency with regard to the AI system provider[161].

The texts recently adopted or in the process of being negotiated could offer new perspectives that would make it possible to cover the intermediation function specific to voice assistants. These reforms are particularly justified because of the central place that this interface now occupies within ecosystems, as a true "armed wing" of the structuring actors.

---

[160] *Business and Professions Code*, Division 7, Part 3, Chapter 6, Section 17940.
[161] For a critical analysis of this minimal framework, see *supra*.

II - VOICE ASSISTANTS AND MARKET ACCESS

In 2020, the Commission launched a sector inquiry into the consumer Internet of Things in the European Union. The reasons for this inquiry included the rapid expansion of this sector in recent years and the projected doubling of voice assistants by 2024. The report states that consumers are wary of the risks induced by the use of these technologies on the respect of their privacy, the protection of personal data, interoperability and lack of compatibility with other devices or systems[162]. The report also establishes that the market's leading actors leave little room for innovation carried by alternative initiatives that are struggling to penetrate a technically expensive but attractive market due to human-machine interaction through voice. The Commission was therefore interested in understanding the connected object environment and in particular the voice assistant environment, the competitive landscape and the trends in this developing sector[163].

In fact, the network effect is particularly developed in this market in that it considerably influences the negotiations between the various actors in a market deployed as an ecosystem.

The network effect. The competitive stakes are high because the network effects are undeniable (brand reputation, confidentiality, the quality of the applications offered and the number of users) and particularly developed on the scale of the voice assistant. This is also what the US House of Representatives points out considering that "*Market participants suggest there are several barriers to entry to compete with general voice assistant platforms. These include overcoming the network effects early entrants have benefited from, including financial investment in hardware, software, and infrastructure, and the ability to sell voice assistant-enabled devices at a discount. Like many platform-based businesses, the voice assistant market benefits from network effects. The more users on a platform, the more third-party devices and applications become available, which attracts more users to the platform. These network effects for voice assistant platforms are amplified by machine learning and artificial intelligence (AI). Improvements in Natural Language Processing (NLP) and AI are expected to improve the quality of voice assistants and contribute to wider adoption*"[164].

Contractual issues. The structuring actors occupy strategic positions on the market in the form of a powerful oligopoly[165]. Above all, they are able to make user companies adhere to their environment, without demanding exclusivity from them, otherwise they

---

[162] European Commission, Final report - Sector inquiry into consumer Internet of Things, 20 January 2022, points 3 and following.
[163] Idem, point 10.
[164] United States House of Representatives, *Investigation of competition in Digital Markets*, Majority staff reports and recommendations, Part I, July 2022, p. 102-103.
[165] On this qualification with regard to virtual assistants, see V. Noskova, "Virtual assistants as gatekeepers for consumption? - how information intermediaries shape competition", *European Competition Journal* 2022, DOI: 10.1080/17441056.2022.2129771.

would be reduced to the rank of invisibles. They have the ability to filter content, which is due to the nature of the voice assistant as a result engine. This helps to significantly increase the bargaining power of service distributors (such as manufacturers of connected speakers and other connected objects that extend the ecosystem or providers of voice assistants). These play a major role in the functioning of the ecosystem because *"they are in a position to capture a significant part of the value and to impose the conditions of provision of this value"*[166]. Similarly, technical solutions providers may have significant bargaining power, particularly with regard to content publishers, except for the most attractive ones, which has an influence on the user's choice of equipment. In most cases, the others are strangled in the negotiation. The power of compulsion is strong when the content publisher wants to be accessible on one of the widely adopted voice assistants to reach a larger audience. What is imposed on them is the cost of developing voice applications - *skills*, actions, competences -, the cost of access to the catalogue and the technical effort made by start-ups and other SMEs to integrate the system... with the promise of mutual and efficient collaboration. In reality, the informational asymmetry is blatant due to a power asymmetry. The Commission, as part of its sector inquiry into consumer connected objects, draws attention to the fact that *" some companies' agreements contain clauses that amplify commercial imbalances between them and the weaker contractual party"*[167]. This imbalance is consistent with known issues relating to application stores, which have given rise to several actions under competition law[168] as well as various convictions for restrictive competitive practices, notably in France following actions by the Minister of the Economy[169]. On the basis of the prohibition of significant imbalance set out in Article L. 442-6, I, 2° of the French Commercial Code, the Commercial Court of Paris ruled that several clauses contained in the contracts concluded between Google LLC, Google Ireland and Google France and application developers were unlawful, including those relating to price, liability waivers, or those allowing for modification of the contract and its unilateral suspension, as well as those producing asymmetrical termination conditions[170].

Finally, the accumulation of powerful intermediaries between content publishers and users raises questions about the capture and distribution of value for publishers who are at the end of the contractual chain. Within the framework of the HADOPI-CSA study dedicated to voice assistants and connected speakers, the creation of a French or European aggregator is envisaged to *"rebalance the balance of power between local and international actors"*[171]. The availability of cultural content depends on it.

---

[166] HADOPI / CSA, *Assistants vocaux et enceintes connectées*, 28 May 2019, pp. 65-66.
[167] European Commission, Final report - Sector inquiry into consumer Internet of Things, 20 January 2022, point 489.
[168] See in particular the pending proceedings before the European Commission against Apple regarding the App Store rules applicable to streaming music providers. The Commission raises two points in particular: first, the obligation to use Apple's own integrated purchase mechanism imposed on developers to distribute their music streaming application via the App Store; second, the fact that Apple applies certain restrictions to developers preventing them from informing users of iPhones and iPads of other, less expensive, purchase options.
[169] On the cumulative application of this text with the DMA once it has entered into force, see M. Behar-Touchais, "L'articulation du DMA avec les autres textes", *Concurrence* No.3-2022, p. 51.
[170] T. com Paris, 28 March 2022, No.2018017655: J.-C. Roda, *CCE* 2022, study No.12, No.7 and R. Amaro, *Concurrences* No. 2-2022, Art. No. 106239- see already T. com. Paris, 2 September 2019, Amazon, No. 2017050625.
[171] HADOPI / CSA, Study supra, p.66.

Effectively, cooperation only exists between the market's structuring actors. Sometimes the contractual relationship gives rise to bitter negotiations, sometimes the battle is won outright by the voice assistant provider once it has locked down its operating system to provide the user only with its own products or services and pre-installed applications, which is essentially the case with Siri. Since voice assistants make cultural content available to users, one wonders about the importance of cooperation between technology providers, publishers and rights holders, especially in a globalized context. So, contractual negotiations are global, which multiplies both the economic actors more or less present on the market and the number of contractual negotiations to be conducted. In this regard, the Commission notes that "*many consumer IoT service providers seem to be unable to negotiate discoverability conditions with providers of consumer IoT technology platforms, which only make exceptions to their general contractual terms and conditions for large counterparties with significant leverage to negotiate. Some stakeholders fear that voice assistant providers would – going forward - agree to carry only those consumer IoT services in exchange for payment, thus affecting the profitability of consumer IoT service providers*"[172].

"Colonization" of ecosystems. The interdependent links between the voice assistant, the core platform services and the user companies, in the sense of the DMA, allow it to "*colonize the ecosystems*"[173], starting with its own before impacting those of the other structuring actors in the market. As such, the voice assistant is a major piece of the ecosystem embodying a new functionality that allows the user to easily penetrate it. This extension of the market is of concern to the European Commission, which has launched several initiatives to clarify the definition of the market and adapt it to digital technology. On 8 November 2022, it published a revised draft Commission notice designed to update a twenty-five year old text[174] that delineates the perimeter within which a company's competition takes place. The Commission proposes new guidance on market definition for digital markets, such as multi-faceted markets and "digital ecosystems" with respect to products built around a mobile operating system. It also provides more guidance on geographic market definition for a text scheduled for late 2023. However, it is not so much the definition of the market that is impacted as the legal consequences, making it necessary to adopt an *ex ante* regulatory approach in addition to the mainly *ex post* competition law. Indeed, the voice assistant appears as a new function on the market, a new form of intermediation, offering the operator the ability to impose its bargaining power and to promote its own competing services by restricting the user's access to the device's functionalities through a voice-only interface. The user is therefore not preserved from its effects. He has priority access to services and products highlighted by the brand that can strongly identify the voice assistant. He is poorly or even wrongly informed, subjected to misleading

---

[172] European Commission, Final report - Sector inquiry supra, point 482.
[173] Here we cite the expression of J.-R. Roda, "Dépasser les limites de l'antitrust", *D.* 2023 (forthcoming) - see also M. E. Stucke, *Breaking Away. How to Regain control over our Data, Privacy and Autonomy*, Oxford University press, 2022.
[174] European Commission, *Notice on the definition of the relevant market for the purposes of Community competition law* (97/C 372/03), 9 December 1997.

interfaces, pre-installed applications by default, and a lack of visibility of certain content, which can hinder access to culture and lock in his choices. The single answer implies a real restriction of the exposure of works and services, and even of access to the market.

However, these objects are intended to take a place in the daily life of users, even intimately so, and to become their active day-to-day assistant, in other words, a companion, a butler, or even a clone taking the features of an avatar. But, at the same time, the voice assistant acts as the "armed wing" of the ecosystem. Voice assistants listen to the user in the literal and figurative sense of the word, operating within closed and constrained spaces such as the home and the vehicle, able to capture the voice and emotions in all circumstances to respond in the most appropriate way. This is the intended effect, as the House of Representatives inquiry notes, " *Voice assistant technology improves at a faster rate when there are more users providing the voice samples needed to train AI. In testimony to the Subcommittee, Professors Maurice Stucke and Ariel Ezrachi describe this a "Learning-by-Doing." As they note : Learning–by–doing network effect is not limited to online searches, but will be present in any environment in which algorithms evolve and adapt based on experience, such, for example, the development of voice recognition or other instances based on machine learning*"[175].

Market access: the issues at stake. This ecosystem, which is extremely intrusive in its purpose, must therefore be covered at a very early stage in the process of entering the market so that it is able to function properly. It is in this perspective that the American Innovation and Choice Online Act in the United States aims to encourage interoperability and prohibit discriminatory and inequitable self-preference. The European legislator went further, recognizing that the voice assistant must not position itself as a tool to "colonize" ecosystems to stifle the bargaining power of its partners, block the consumer in its ecosystem or "nudge" the consumer towards it. While the law of anti-competitive and restrictive practices remains applicable[176] but does not prevent such behaviour, it is therefore appropriate to impose *a priori* obligations on voice assistants to ensure the fairness and contestability of digital markets. The DMA has taken on these issues through its cross-sectoral approach. To this end, it provides for families of obligations on gatekeepers, from which we can distinguish the general obligations set out in Article 5, supplemented by a series of specific obligations that are scattered throughout the text.

As a result, gatekeepers are encouraged to stop requiring their operating system to be installed by default. A multi-choice screen will have to be proposed to users so they can opt for a competing service (Article 6.3). They will no longer be able to promote their services and products in relation to those of the sellers who use their platform (Article 6.5) or to exploit the data of the sellers to compete with them (Article 6.2), nor

---

[175] House of Representatives, Inquiry supra, p. 103.
[176] On this question, see M. Behar-Touchais, "L'articulation du DMA avec les autres textes existants", *supra.*

will they be able to impose on the app developers certain additional services such as the payment system (Article 6.4). If virtual assistants are qualified as core platform services in the sense of the DMA, obligations may be imposed on them such as the obligation of free interoperability (Article 6.7), the prohibition of combining users' personal data from other services (Article 5.2), the obligation to communicate their figures to advertisers and publishers (Article 5.9 and 10), or the obligation to allow dynamic and real time data portability (Article 6.9) Article 14 also provides for an obligation to inform about concentrations. These services may also be subject to market investigation under Chapter IV. It should be noted that the obligations set out in Article 5 of the DMA are cross-sectoral, in that they apply to all companies designated as gatekeepers, whereas the list of obligations set out in Article 6 will allow the Commission to impose them on a case-by-case basis on these operators.

Regulation 2022/1925, Article 5: Obligations for gatekeepers

1.   The gatekeeper shall comply with all obligations set out in this Article with respect to each of its core platform services listed in the designation decision pursuant to Article 3(9).

2.   The gatekeeper shall not do any of the following:

(a) process, for the purpose of providing online advertising services, personal data of end users using services of third parties that make use of core platform services of the gatekeeper;

(b) combine personal data from the relevant core platform service with personal data from any further core platform services or from any other services provided by the gatekeeper or with personal data from third-party services;

(c) cross-use personal data from the relevant core platform service in other services provided separately by the gatekeeper, including other core platform services, and vice versa; and

(d) sign in end users to other services of the gatekeeper in order to combine personal data, unless the end user has been presented with the specific choice and has given consent within the meaning of Article 4, point (11), and Article 7 of Regulation (EU) 2016/679.

Where the consent given for the purposes of the first subparagraph has been refused or withdrawn by the end user, the gatekeeper shall not repeat its request for consent for the same purpose more than once within a period of one year.

This paragraph is without prejudice to the possibility for the gatekeeper to rely on Article 6(1), points (c), (d) and (e) of Regulation (EU) 2016/679, where applicable.

3. The gatekeeper shall not prevent business users from offering the same products or services to end users through third-party online intermediation services or through their own direct online sales channel at prices or conditions that are different from those offered through the online intermediation services of the gatekeeper.

4. The gatekeeper shall allow business users, free of charge, to communicate and promote offers, including under different conditions, to end users acquired via its core platform service or through other channels, and to conclude contracts with those end users, regardless of whether, for that purpose, they use the core platform services of the gatekeeper.

5. The gatekeeper shall allow end users to access and use, through its core platform services, content, subscriptions, features or other items, by using the software application of a business user, including where those end users acquired such items from the relevant business user without using the core platform services of the gatekeeper.

6. The gatekeeper shall not directly or indirectly prevent or restrict business users or end users from raising any issue of non-compliance with the relevant Union or national law by the gatekeeper with any relevant public authority, including national courts, related to any practice of the gatekeeper. This is without prejudice to the right of business users and gatekeepers to lay down in their agreements the terms of use of lawful complaints-handling mechanisms.

7. The gatekeeper shall not require end users to use, or business users to use, to offer, or to interoperate with, an identification service, a web browser engine or a payment service, or technical services that support the provision of payment services, such as payment systems for in-app purchases, of that gatekeeper in the context of services provided by the business users using that gatekeeper's core platform services.

8. The gatekeeper shall not require business users or end users to subscribe to, or register with, any further core platform services listed in the designation decision pursuant to Article 3(9) or which meet the thresholds in Article 3(2), point (b), as a

condition for being able to use, access, sign up for or registering with any of that gatekeeper's core platform services listed pursuant to that Article.

9. The gatekeeper shall provide each advertiser to which it supplies online advertising services, or third parties authorised by advertisers, upon the advertiser's request, with information on a daily basis free of charge, concerning each advertisement placed by the advertiser, regarding:

(a) the price and fees paid by that advertiser, including any deductions and surcharges, for each of the relevant online advertising services provided by the gatekeeper,

(b) the remuneration received by the publisher, including any deductions and surcharges, subject to the publisher's consent; and

(c) the metrics on which each of the prices, fees and remunerations are calculated.

In the event that a publisher does not consent to the sharing of information regarding the remuneration received, as referred to in point (b) of the first subparagraph, the gatekeeper shall provide each advertiser free of charge with information concerning the daily average remuneration received by that publisher, including any deductions and surcharges, for the relevant advertisements.

10. The gatekeeper shall provide each publisher to which it supplies online advertising services, or third parties authorised by publishers, upon the publisher's request, with free of charge information on a daily basis, concerning each advertisement displayed on the publisher's inventory, regarding:

(a) the remuneration received and the fees paid by that publisher, including any deductions and surcharges, for each of the relevant online advertising services provided by the gatekeeper;

(b) the price paid by the advertiser, including any deductions and surcharges, subject to the advertiser's consent; and

(c) the metrics on which each of the prices and remunerations are calculated.

In the event an advertiser does not consent to the sharing of information, the gatekeeper shall provide each publisher free of charge with information concerning the daily average price paid by that advertiser, including any deductions and surcharges, for the relevant advertisements.

In order to understand all the issues related to the voice assistant market, it's important to examine more specifically the key points that will allow non-discriminatory and equitable access to the market in order to guarantee the user's freedom of choice, particularly when accessing cultural content. Thus, it is appropriate to question self-preference in all its facets (A), the interoperability of systems and applications (B), and access to technical and user data (C).

A - Self-preference (not blocking access to the market)

Self-preference: a major issue. For voice assistant providers, one of the main concerns is the potential for self-preference which, as discussed above, can arise from two dimensions. First, the voice assistant operator's speaker is preferably used to interact with the voice assistant. Then, the preferred applications to answer the user's requests will be those of the same operator. Finally, the lack of a screen in the interaction with voice assistants as well as the challenges associated with the single answer on these systems may lead to favouring the services of these operators by default.

The final report of the European Commission's IoT inquiry notes the many concerns of stakeholders in this regard. These include pre-installation of the voice assistant and applications, vendor-defined defaults, and highlighting of products and services. For example, when it comes to "out-of-the-box" functionality offered to users, respondents indicate that such practices *"impact the discoverability of consumer IoT services to the extent that it becomes a competitive (dis-)advantage"*[177] and that large, globally active creative content service providers would particularly benefit. In this respect, it is emphasized that pre-installation practices are widespread - the voice assistant is usually programmed to use a specific service by default to perform a specific activity - which most often lead voice assistant providers to direct the user to their own related services (e.g. music, podcasts, books, audiovisual content, shopping). Some respondents then explain that "*pre-installation constitutes a competitive disadvantage for consumer IoT service providers whose applications are not pre- installed, as a user has to take additional steps to access a service that is not included in the "out of the box" offer. In contrast, pre-installed services profit from higher levels of user discoverability*"[178]; "*It is also alleged that defaults are 'stickier', meaning that they result in users staying with the default service for longer*"[179] *and that accessing another service provider will involve adding a specific invocation name to one's voice command (*e.g., *"play song X on service Y)*". The effects of prominence practices determining the

---

[177] European Commission, *Final report - Sector inquiry into consumer Internet of Things,* 20 January 2022, point 440.
[178] European Commission, Final report - Sector inquiry supra, point 449.
[179] European Commission, Final report - Sector inquiry supra, point 458.

visibility and findability of a service are also considered. These practices are subject to significant negotiation[180].

The EU Commission's report further highlights the existence of exclusivity and tying practices for voice assistants, as well as practices that limit the ability to use different voice assistants on the same smart device[181]. In particular, it cites the case of voice assistant exclusivity on a smart device. The results of the inquiry reveal that the manufacturers of these objects generally offer users devices with only one integrated voice assistant, whereas there is a demand on the part of users for dual control of the device by a generalist voice assistant (the most common situation) and a specialized voice assistant (which is provided by the competition). To this end, the Commission specifies that "*simultaneous use of voice assistants, i.e., switching from one voice assistant to another using a specific activation word to activate one of the voice assistants, is only possible on a limited number of smart devices manufactured by the respondents. Some concerns about the inability to enable simultaneous use of voice assistants on smart devices were raised in the industry inquiry*". As it stands, this is one of the concerns of actors who position themselves in the market as a simple voice interface. However, the specific activation word is linked to brand identification sometimes at the expense of the brand of the connected objects triggered by the voice assistant command. As a result, these practices can have the effect of reinforcing the identity of the voice assistant in the home environment. Audiovisual services have similar or even additional concerns. In addition to confining the user in an ecosystem, this ultimately raises the question of the future of free and universal terrestrial television and radio, depending on the evolution of voice interfaces and technologies used if the voice assistant were to become the command centre for audiovisual services.

The report notes that there are also concerns about the risk of disintermediation[182]. One such risk lies in the position of voice assistants and smart device operating systems as intermediaries between users and smart devices or IoT services for consumers insofar as this position, coupled with their key role in data generation and collection, would allow them to control relationships with users. In this context, stakeholders also expressed concerns about the discoverability and visibility of their IoT services to consumers. They then fear a loss of brand recognition and direct relationship with users, as well as "*lack of automatic data gathering prevents third-party consumer IoT service providers from customising the user experience in real time and reportedly puts them at a competitive disadvantage with respect to the first-party services offered by leading general-purpose voice assistant providers*"[183].

---

[180] European Commission, Final report - Sector inquiry supra, points 464 and following.
[181] European Commission, Final report - Sector inquiry supra, points 470 and following.
[182] European Commission, Final report - Sector inquiry supra, points 474 and following.
[183] European Commission, Final report - Sector inquiry supra, point 483.

Self-preferential practice(s). Different procedures for self-preference can thus be identified. Strictly speaking, self-preference will refer to self-referencing - or self-preference in ranking - which is aimed at the case where the voice assistant gives priority to related products or services offered by the company providing the voice assistant. In the broad sense, self-preference can capture cases of bundled sales, i.e. when a set of goods and services are provided in a single package. In the case of voice assistants, this practice may consist of pre-installing an assistant on the user's hardware (smartphone, computer, speaker), or pre-installing various voice applications on the voice assistant when it is delivered to the user. This practice can be reinforced by setting default values that consist, for example, of favouring the launch of the application located in the proprietary ecosystem of the voice assistant provider when the user requests it. Self-preference can also result from practices that exclude competitors or disadvantage them in favour of the related services of the company providing the voice assistant. These practices may include pricing or data processing, insofar as the platform operator relies on the data of its business users to improve its own offerings to compete as a service provider or third-party vendor.

Self-preference sanctioned by competition law. Several actions have been taken on the basis of competition law specifically targeting the various self-preference practices of the major digital operators. In addition to the Microsoft litigation[184], it is worth mentioning in this respect the conviction of Google in the Google Shopping case for having artificially manipulated the algorithm of its search engine in order to give preference to its own comparator, which confirms that self-preference can, depending on the circumstances, constitute an abuse of a dominant position[185]; also in the Google Android case, this time concerning the pre-installation of browsers and search engines contractually imposed on suppliers of mobile devices[186]. In addition, an investigation has been opened regarding the obligation imposed on suppliers of Android-based devices to pre-install Google Assistant to the exclusion of the service provided by its competitors[187]. The European Commission has also "*opened new fronts*"[188] by specifically targeting Amazon and Apple's practices. Regarding Amazon, a first statement of objections was sent by the Commission concerning the use of non-public data of professional users of its marketplace in order to favour its own products[189]. In addition, an investigation was launched into the "buy box" and Amazon Prime in order to analyse the practice of self-preference, which consists of giving preference to one's own retail activity or to those of distributors using the marketplace's logistics services[190]. In July 2022 - on the eve of the DMA's adoption - Amazon made a proposed commitment to the European Commission that addressed, among other

---

[184] CFI, 17 September 2007, Microsoft Corp. v. Commission, case T-201/04.
[185] Trib. EU, 10 November 2021, Google Alphabet v. Commission, case T-612/17.
[186] Trib. EU, 14 September 2022, Google Alphabet v. Commission, case T-604/18.
[187] F. Y. Chee and N. Balu, "Google's Voice assistant in new EU antitrust investigation," *Reuters*, 9 September 2021.
[188] J.-C. Roda, "Un an de droit de la concurrence dans l'univers numérique", *CCE* 2022, study No.12, No.5, which then describes these various disputes.
[189] Case COMP/AT.40462.
[190] Case COMP/AT.40703.

things, the Commission's concern about self-referral when assigning the buy box[191]. In addition, the Commission issued a statement of objections against Apple for requiring application providers to use Apple's integrated purchasing mechanism in order to distribute their music streaming application. The Commission also noted that the operator limited the ability of app providers to use alternative in-app payment providers and forced them to use its own system (Apple Pay) while requiring them to pay commissions that its own Apple Play music subscription service was not subject to. This resulted in a lack of attractiveness of third-party applications[192].

Self-preference framed by the DMA: do's and don'ts. In the wake of these various disputes and in order to guarantee the contestability and fairness of digital markets, the DMA is coming to grips with these self-preferential practices in various ways, following the "do's" and "don'ts" method characterizing the text[193]. It is particularly interested in the mechanisms for ranking the products and services made available to the user as part of the single response, but also in the practices of installing features and applications by default.

As a preventive measure, the text first prohibits gatekeepers from practising self-referencing, in the sense of self-preference in the ranking, by expressly targeting virtual assistants. For example, Article 6.5 provides that "*The gatekeeper shall not treat more favourably, in ranking and related indexing and crawling, services and products offered by the gatekeeper itself than similar services or products of a third party. The gatekeeper shall apply transparent, fair and non-discriminatory conditions to such ranking*"[194]. In addition, the so-called anti-circumvention rule of Article 13 of the DMA may lead to the prohibition of self-referencing and other distortions of recommendations that tend to impose a certain behaviour on users of the gatekeeper's service, in that this provision stipulates that "*The gatekeeper shall not degrade the conditions or quality of any of the core platform services provided to business users or end users who avail themselves of the rights or choices laid down in Articles 5, 6 and 7, or make the exercise of those rights or choices unduly difficult, including by offering choices to the end-user in a non-neutral manner, or by subverting end users' or business users' autonomy, decision-making, or free choice via the structure, design, function or manner of operation of a user interface or a part thereof*"[195]. During the negotiations of the text, some may have advocated that the prohibition on self-preference "*be extended beyond the ranking of offers to any technique for influencing end-users toward the products and services of the gatekeeper and related*

[191] https://ec.europa.eu/competition/antitrust/cases1/202229/AT_40462_8414012_7971_3.pdf
[192] Case COMP/AT.40437.
[193] L. Idot, "Propos introductifs", *Concurrence* No.3-2020, p. 45, No.20.
[194] Ranking is defined by Article 2.22 of the DMA as "*the relative prominence given to goods or services offered through online intermediation services, online social networking services, video-sharing platform services or virtual assistants, or the relevance given to search results by online search engines, as presented, organised or communicated by the undertakings providing online intermediation services, online social networking services, video-sharing platform services, virtual assistants or online search engines, irrespective of the technological means used for such presentation, organisation or communication and irrespective of whether only one result is presented or communicated*".
[195] In this sense, see M. Peitz, *The prohibition of self-preferencing in the DMA*, Issue Paper, CERRE, November 2022, p. 10.

companies"[196]*.* This at least appears to be the case when the DMA, in Article 6.2, limits the use of data by the gatekeeper in competition with the user companies of its service; it then refers to "*any data that is not publicly available that is generated or provided by those business users in the context of their use of the relevant core platform services or of the services provided together with, or in support of, the relevant core platform services, including data generated or provided by the customers of those business users*". The same applies to the prohibition of misleading or manipulative choice interfaces - or *dark patterns* - set out in Article 13, which will apply beyond ranking alone (see below). Beyond these prohibitions, the text addresses a number of other practices, but this time in the form of positive obligations, which consist of requiring gatekeepers to limit the lock-in of their users through the mechanism of pre-installation and default values to guarantee their freedom of choice (see below); this also justifies imposing interoperability measures and guaranteeing access to data.

In reality, self-preference practices are the major risk that must be addressed as a priority in order to ensure the fluidity of the market by setting up the right prevention and protection barriers. This approach is particularly complex, as it involves capturing all emerging practices in the marketplace in order to evolve the DMA's preventive approach. The objective of the text is to open up ecosystems while ensuring their equitable operation, by identifying a few *a priori* practices that contribute to locking them in. By its preventive and cross-sectoral aim, the DMA approach will tackle practices that are capable of structurally unbalancing the ecosystem. These practices cannot be permanently fixed in a list of obligations. To this end, the DMA provides that the obligations set out in Article 6 will evolve in accordance with the mechanisms set out in Article 8. This preventive approach will be accompanied by a second level of *ex post* intervention to sanction, this time on the basis of competition law, the cyclical imbalances in the market that have slipped through the cracks of the first level of regulation.

B - Interoperability (to access the market)

Why interoperability? The cost of technology investment, the large (mostly indirect) network effects, and the many other lock-in effects at work (e.g., the inability of users to retrieve their data) are undeniable barriers to entry or expansion in the market for general-purpose voice assistants in the near term. In this respect, it should be noted that, in addition to economies of scale and scope (covered by the mention of the necessary investments), network effects (when they are captured by the actors in place) constitute classic sources of market failure in industrial economics. Particular attention must be paid here to indirect network effects since they are particularly powerful with regard to voice assistants. Indeed, a large share of end users increases the value of the service for application providers (and vice versa). This makes the few

---

[196] Senate, *Résolution européenne sur la proposition de Règlement sur les marchés numériques* (DMA), No.32, 12 November 2021, point 49.

voice assistants on the market unavoidable for both groups of users; especially since it is complicated for users to use multiple voice assistants (see above). In this situation, interoperability helps limit the ability of the few voice assistant vendors to unduly control access to their must-have infrastructure.

In its investigation of digital markets, the House of Representatives illustrates the phenomenon as follows:

" *As the voice assistant market expands, it may be difficult for users to switch between platforms. Because voice assistant platforms are not always interoperable, users would incur costs to purchase one or more new devices. Moreover, voice assistant technology is designed to learn its user's preferences over time. These preferences range from settings like billing information and default services for responding to music commands to more advanced learning like past voice commands and shopping history. As a voice assistant improves its "understanding" of its user, it may increase the costs associated with switching to another platform. As one market participant noted in a submission to the Subcommittee, "the user may become more dependent on that particular voice assistant and be far less likely to use a rival voice assistant that has not yet 'caught up' with the user's preferences."*

The design of most voice assistants—specifically on screenless devices—amplifies the ability of voice assistant platforms to favor their services as a default or as a response with limited choice.667 This dynamic makes it easier for popular voice assistants to favor their first-party services.

There is also a significant potential for misuse of data to harm competition or consumers.
Similar to other platforms, such as cloud and operating systems, voice assistant platforms collect and store users' interactions with the voice assistant. During the investigation, several companies shared concerns that voice assistant platforms would be able to use this vantage to glean competitive insights from third-party voice applications or smart appliances that are performing well. As a result, platforms could use that data to acquire competitive threats or integrate their features into the company's product.

Privacy and data experts have also commented that the smart home ecosystem is some of the most sensitive data that can be collected. Voice assistant platforms not only record voice interactions, but also receive information about the skills used—"whether a light is on or off. Or, if a customer links Alexa to a third-party calendar skill, Alexa may receive information about the events on the customer's calendar."670 This raises significant concerns regarding whether a person has provided consent to data collection. Voice assistants not only collect information on the primary user, but also people in their environment, including children. Finally, leaders in the voice assistant ecosystem set the rules for third parties. To make a voice assistant enabled device, market participants must comply with voice assistant platform vendor specifications. As Mr. Spence of Sonos noted in his testimony before the Subcommittee:

To gain access to their platforms and integrate with their services, these companies issue all manner of take-it-or-leave-it demands, from early and technically detailed access to our product roadmaps, to proprietary business data, including sales forecasts, to waivers of essential contractual rights.

The Subcommittee also heard from multiple voice assistant developers that have struggled to gain access to key functionality needed to build their applications, such as the unprocessed user

commands.672 While still developing, the voice assistant market shows early signs of market concentration"[197].

Definition of interoperability. The term "interoperability" itself covers multiple realities, but in general, *"it is the ability of a system product, whose interfaces are fully known, to work with other existing or future products or systems without restriction of access or implementation"*[198], without any particular effort from the user. In the context of a simplified and global approach to interoperability, PEReN defines it under three aspects: interoperability itself, covering its horizontal or vertical form, the compatibility of systems and the portability of data. All of this can be found in a very gradual way within the ecosystems, ranging from a total absence of interoperability, to complete interoperability, to interoperability concerning only certain functionalities shared horizontally by all actors[199].

PEReN: Definition of interoperability

"**Horizontal interoperability** (between competing products, such as between social media)[200] and **vertical interoperability** (interoperability with complementary products, such as between a marketplace and third-party tools used by sellers on the platform)."
"**Compatibility** is a limited form of interoperability, which allows two specific but different types of systems to communicate with each other."
"**Data portability**[201], on the other hand, is only about the migration of data between two services or two platforms".

Source: PEReN, Shedding light on…N°1 // Interoperability, October 2021, p.1.

Interoperability: what objectives? The performance of the voice assistant depends on the applications from which it will propose a response to the user's command. However, the development of new products and services is hampered by the vertical integration of major suppliers, who combine their own products and services with those

---

[197] House of Representatives, *Investigation of competition in Digital Markets*, majority staff reports and recommendations, Part I, July 2022, p. 104 and following.
[198] French association of free software users cited by PEReN, *Shedding light on…N°1 // Interoperability*, October 2021, p.1 - see also Interministerial Directorate for the Digital Economy and the State Information and Communication System, "General Interoperability Framework: Standardize, Align, and Focus to Exchange Effectively," Version 2.0, 2015 - comp. article 2 DMA defining interoperability as *"the ability to exchange information and to mutually use the information which has been exchanged through interfaces or other solutions, so that all elements of hardware or software work with other hardware and software and with users in all the ways in which they are intended to function"*.
[199] PEReN, Report supra.
[200] Horizontal interoperability is more of a remedy for problems related to direct network effects, which does not seem totally relevant in this context (it does not require modification, horizontal interoperability is not mentioned too much).
[201] Beyond network effects, dynamic portability (which is a form of interoperability) as provided for in the DMA (Article 6.9) could also be of interest. This remedy is not related to network effects but more to other lock-in effects, such as the impossibility of retrieving one's data (in real time) to join another service.

of third parties to provide an attractive offering. Therefore, the ability to interconnect the different components of an ecosystem within the framework of interdependent links is a particular question. The various components of the ecosystem must be able to communicate with each other, which means ensuring interoperability between connected devices, voice assistants and the services offered to users. Without interoperability, it is impossible to offer users a variety of products and services, thus preventing self-preference. In addition to application providers, it would also be interesting to add that interoperability allows end-users, on the other side of the platform, to benefit from indirect network effects (i.e. a wider range of services on the platform, thus more choice, and in theory a positive effect on quality, price, and even security and privacy if this is part of their choice criteria). Moreover, interoperability allows new entrants to a market to take advantage of the "network effect" that exists through the interplay of structuring actors, to use it to their own advantage and to move up the market in their turn[202]. This is usually due to the availability of a protocol or standard.

In this respect, interoperability is the best lever for asymmetrical *ex ante* regulation[203], but it is only one tool among others, whose impact must be evaluated in a contextual manner, both in terms of its economic issues (leverage effect, cost of entry into the market, economic model), technical issues (difficulties in opening interfaces, data security) and legal issues (competitive and contractual aspects, data portability, protection of intellectual property rights).

Observation of the lack of interoperability between voice assistants, and between voice assistants and third-party user companies. This is the finding of the European Commission, which, in its sectoral inquiry on the Internet of Things for the general public[204], focuses on interoperability issues. This is the technical element that determines the creation of ecosystems and the main lever for opening them up with a view to guaranteeing market access. The Commission identifies two nodes blocking market access. Firstly, the system operator controls the possibility of market integration through proprietary technical standards, and secondly it will control the functionality of voice assistants to limit access.

Voice assistants arrived on the market before the definition of a standard or a communication protocol was considered. The dominance of proprietary technologies has sometimes led to the creation of "de facto" standards, accompanied by certification processes. These few structuring actors unilaterally control the interoperability and integration processes and are able to limit their functionalities. The technical specifications and the software specifications to enable access to voice assistants are made available to third parties subject to the conclusion of affiliation agreements, the

---

[202] C. Doctorow, remarks collected by H. Guillaud, "Dé-monopoliser l'Internet par l'interopérabilité", *InternetActu*, 11 October 2021.
[203] PEReN, Report supra, p.1.
[204] European Commission, *Final report - Sector inquiry into consumer Internet of Things*, 20 January 2022.

negotiation of which is reserved for the structuring actors only. These are able to independently define the requirements needed to achieve interoperability with their proprietary technology, imposing terms and conditions, technical requirements and certification processes that they govern unilaterally. In addition, the structuring actors have full control over the experience of their own products and services from the very beginning of the user interaction, collect the user data that is useful to them and, in most cases, are completely autonomous in resolving technical issues, dictating technical performance and processes to third-party user companies. When the voice assistant is configured by default, linked to the operating system, it prevents the installation of another voice assistant on a terminal. In addition, some voice assistants cannot be configured to change search services, so the only sources queried to answer the user's questions are pre-selected by the terminal. Finally, in the specific context of connected speakers, the developers of applications for these speakers cannot interface their services with those of the speaker, but are required to provide applications through the voice assistant, which can only receive them if they are integrated into their system[205].

The result, at the API scale, is that only a few general-purpose voice assistant actors can determine who can work with which services and which devices. Acting on the interoperability of ecosystems avoids market fragmentation leading third-party user companies to develop a plurality of content and functionalities in order to be accessible via several voice assistants, and therefore to be present (or not) on each ecosystem. In theory, this would promote freedom of choice for users (in terms of quality, price or other elements such as the security or confidentiality of the services used) and therefore innovation (new services could more easily be created and could access the necessary platforms so that they can be distributed). However, the cost of deploying functionality and accessibility is so high that it could lead to the business model of the user companies being called into question or disappearing from the market. The focus is on the structuring actors (Apple, Amazon, Google, etc.,) leaving in the shadow thousands of user companies that are struggling to get a foothold in the market. Especially since, for the user, the choice of a voice assistant is generally based on the reputation, the brand. Access to content comes later, and due to the absence or incompleteness of interoperability between systems, the user is quickly trapped in an integrated system that is difficult to set up, and for which unsubscribing remains a complex operation.

Proposals for a solution. The common-sense principle is to focus the bulk of the obligations on the main platforms so that they become interoperable with third-party user companies. In approving the Commission's decision condemning Microsoft for its refusal, constituting an abuse of a dominant position, to provide information relating to the interoperability of the Windows operating system with competing products, the Court of First Instance observed that, if it is established that the information is

---

[205] European Commission, Final report supra, point 380.

indispensable for Microsoft, it is necessarily of great value to competitors who wish to have access to it[206]. To the argument that the disclosure of server-to-server protocols was subject to intellectual property rights, which implied that they had to remain secret in order not to undermine innovation, which Microsoft justified by the free competition exception established by the Magill[207] and IMS Health[208] rulings, the Court replied that there was no objective justification or exceptional circumstances authorizing the refusal to grant a license within the meaning of the above-mentioned case law[209]. As a result, the interoperability requirement is imposed on Microsoft to preserve free competition while respecting users' rights.

Today, the difficulty in this fragmented landscape lies in taking into account pre-existing tools to build interoperability. The data held by the structuring actors must circulate without technical hindrance - which could not be justified by security issues or legally protected interests - and be put at the service of effective competition and user protection. In this respect, in the Nintendo case, the CJEU questioned technical protection measures that could go beyond the objective of fighting the production of infringing copies of video games to circumvent interoperability issues and segment the market. Referring to recital 48 of Directive 2001/29, the Court recalled that "*legal protection against acts not authorised by the rightholder of any copyright must respect the principle of proportionality, (…) and should not prohibit devices or activities which have a commercially significant purpose or use other than to circumvent the technical protection*". Thus, even if it did not take a position on the technical protection measures of the Nintendo system, the CJEU specified that when examining compliance with Article 6 of Directive 2001/29, the national courts, when assessing the principle of proportionality, must take into account the purpose of TPMs and the purpose of the devices enabling their circumvention. The choice of other technical protection measures which could have caused less interference with, or fewer limitations on, the activities of third parties not requiring the authorization of the copyright owner, while providing comparable protection for the rights of that owner, must be preferred[210]. It is also necessary to verify that the technical protection measures are proportionate to the uses.

Therefore, it is important to evaluate the cost of setting up a standard so as not to strengthen the position of the structuring actors and exclude third-party user companies. In this regard, PEReN notes that Amazon's AWS and Microsoft's Azure standards are interchangeable and can be used on competing tools, but because they can modify the APIs unilaterally, user companies remain dependent on these structuring actors. To remedy this situation, the regulation unit recommends that the sudden changes in these APIs be controlled[211]. For its part, the European Commission, in the context of the sector inquiry on consumer IoT, notes the complexity

---

[206] CFI Grand Chamber, 17 September 2007, Microsoft Corp. Against Commission, case T-201/04, point 694.
[207] ECJ, 6 April 1995, RTE and ITP v. Commission ("Magill"), case C-241/91 P and C-242/91 P.
[208] CJEU, IMS Health/NDC Health, case C-418/01, 29 April 2004.
[209] CFI, Microsoft, supra, point 680.
[210] CJEU, 23 January 2014, Nintendo Co. Ltd and others v. PC Box Srl, case C-355/12, point 32.
[211] PEReN, Report supra, p.7.

of the standardization landscape coupled with the equally significant fragmentation of the proprietary technology landscape, affecting potential growth and innovation, by consequently locking users into proprietary ecosystems in the absence of compatibility and interoperability between products, services and technologies from the same provider. It proposes to strengthen the standardization process and simplify it[212]. The DMA introduces obligations that take up these proposals, but adopts a more cross-sectoral approach without going so far as to promote an asymmetrical approach, which could be justified by the position of the structuring actors in each ecosystem.

The vertical interoperability promoted by the DMA. The contestability of markets implies, in the context of an *ex ante* regulatory approach, that obligations be imposed on the structuring actors in order to guarantee the interoperability of platforms.

The general obligations in this area are set out in Article 5 of the text. However, the DMA introduces, more specifically, interoperability obligations that are both horizontal with respect to messaging services (Article 7) but also vertical, including an obligation to access the essential functionalities of operating systems, virtual assistants or the hardware capabilities of a given device that the gatekeepers use for their own products or services (Article 6.7) and the possibility of installing third-party application stores or downloading applications without going through the main player's application store (Article 6.4). Only vertical interoperability obligations are applicable to voice assistants in order to facilitate the entry of third-party user companies giving them access to essential functionalities that they could not reproduce, which may have for some, a leverage effect to expand on the market. Finally, Article 6.9, which will also apply to virtual assistants, provides for the dynamic portability of user data (and therefore interoperability: " *including by providing, free of charge, tools to facilitate the effective exercise of such data portability, and including by the provision of continuous and real-time access to such data. ").* This could theoretically in this context facilitate the transition from one virtual assistant service to another.

Third-party application developers will be able to use all voice assistants at lower costs, choosing at least some that are suitable for their business segments. However, the DMA plays on both sides, considering the restrictions on the user, who must be able to "*install and effectively use third-party software applications or software application stores on hardware or operating systems* (...)" (Recital 50). If such restrictions are not lifted, it is the contestability of the services and products of the gatekeepers that will be lacking by opposing the option for users to change their voice assistant (recital 54). The texts limit the system to competing service or equipment providers who, in order to guarantee the user's freedom of choice, need this access (see recital 55), since they are the key actors in accessing the voice assistant market. This is why the gatekeepers must detail the technical conditions for accessing the operating system (functionalities, security and integrity of the system, modification of

---

[212] European Commission, Final report supra, points 496 and following.

interfaces) and, if necessary, specify the cost of accessing the operating system. It is a question of allowing, downstream, third-party user companies to interconnect with the services and hardware they offer using interfaces similar to those of the gatekeepers. However, the issue of access is broadly understood and involves providing third-party user companies with equivalent conditions to those provided for the gatekeeper. This is bound to introduce difficulties in the absence of a standard. Recital 96 recognizes that the implementation of certain obligations such as interoperability could be facilitated by the use of technical standards. However, even if the development of an open interface standard could be an alternative, it remains cumbersome and complex in the absence of consensus.

It is therefore preferable to let the gatekeepers manage the interfaces by allowing them to guarantee the integrity of their interfaces as provided for in Article 6.7, through the licensing of access seekers[213]. With regard to users, they can take measures to encourage them to protect the security of their applications, on the sole condition that this is neither a pre-installation nor a default setting. Both technical and contractual measures must remain "*strictly necessary and proportionate and (be) duly justified by the gatekeeper*"[214]. They can grant licenses without any objective criteria being set beforehand, as the DMA does not impose any obligation in this regard. However, if they refuse to grant the license, a procedure may be initiated at the European level based on complaints from third-party user companies or if the Commission takes a self-initiated decision. These complaints are moreover complex to define from the outset because of the multiple cases that may arise, especially since the gatekeeper is entitled to take the necessary measures to ensure that the conditions for granting the license do not compromise the integrity and security of the system. Therefore, the granting of an access license based on objective criteria could be considered a necessary and proportionate measure[215]. To ensure innovation and effective competition that preserves the functionality of the voice assistant from economically fragile alternative actors, some note in this regard that the costs of access to services and materials will have to be covered at least partially by third-party user companies[216]. However, the DMA introduces free access in its Article 6, which will inevitably lead to disputes if this provision has a counterproductive effect on the maintenance of competition.

It would probably be necessary to go further in the process by creating a dispute resolution procedure overseen by a regulator with sufficient technical expertise to deal with interoperability issues and guarantee the integrity of the voice assistant, similar to what ARCEP (France's Electronic Communications, Postal and Print media distribution Regulatory Authority) does in the electronic communications context[217].

---

[213] M. Bourreau, *DMA horizontal and vertical interoperability obligations*, CERRE, November 2022, p.14 et seq.
[214] DMA, Recital 50 - see also Articles 6.4, 6.7.
[215] M. Bourreau, *DMA horizontal and vertical interoperability obligations* supra.
[216] *Ibid.*
[217] See also on this point, BEREC, "Opinion on the European Commission's proposal for a Digital Markets Act", BoR (21) 35, 11.03.2021, page 5: "*In this line, BEREC considers that it is essential to include a dispute resolution mechanism in the DMA proposal*".

Article 8 of the DMA provides for this purpose that specific obligations may be specified, if necessary, in delegated acts.

C - Access to data (to open the market)

The voice assistant market is dominated by a small number of very powerful providers who control both access to the services and products offered to users, but also to the data linked to them. The overall expansion of the databases of the structuring actors (1) generates legitimate concerns on the part of all the new entrants about the recovery of data useful to application developers in order to enhance the content offer, or to publishers in order to adapt to the users' profile (2). This data recovery can only be achieved if interoperability is guaranteed, since this is one of the levers to be implemented (see above) in the context of data portability (3).

1. Overall expansion of the databases of structuring actors

Quantity of data. In a Digital markets Inquiry updated in July 2022, the US House of Representatives identifies massive data collection as the main issue in the voice assistant market. It is about training and validating algorithmic systems: "*The scale of users generating data is arguably the most important asset in terms of AI. The incumbents have access to large data sets that—when combined with machine learning and AI— position them to benefit from economies of scope in the smart home*"[218]. In the context of the voice assistant, the quantity of data makes it possible to improve the performance of speech interaction algorithms (Speech-to-Text and Text-to-Speech) and automatic natural language processing.

Cloud storage. Next, the inquiry report looks back at the significant investments made by the leading providers to build the voice assistant universe, both to acquire the physical infrastructure but also to develop the software features as part of a cloud infrastructure: "*Competing as a voice assistant platform also requires significant financial resources. A firm must make significant investments to design and train a voice assistant, as well as acquiring the physical infrastructure: hardware and cloud computing. Additionally, incumbents have also acquired various firms that specialize in voice recognition and natural language processing, a functionality that is used in their voice assistants. For example, both Apple and Amazon acquired companies to develop their core voice recognition technologies, and every incumbent has*

---

[218] House of Representatives, *Investigation of competition in Digital Markets*, majority staff reports and recommendations, Part I, July 2022, p. 103.

continually invested in AI startups to improve their voice assistant ecosystem. Currently, voice assistant software is built on cloud computing infrastructure. In the case of Amazon Alexa and Google Assistant, the voice assistant platforms also own the underlying cloud infrastructure, AWS, and GCP, respectively. Market participants note that advancements in voice assistant ecosystems are beginning to rely on edge computing technology, which brings the computation and data storage closer to the device and is a technology in which the incumbent cloud market leaders have a head start"[219].

Massive collection. Finally, the report emphasizes that the physical investments made are not cost-effective and the objective is not to make them so. They are being deployed on a large scale and on attractive terms to prioritize the massive collection of data from users: "*Market participants have also raised concerns about incumbent firms offering voice-enabled hardware—specifically hubs such as smart speakers—to both collect large amounts of personal user data and strengthen other lines of business. At the Subcommittee's field hearing, Sonos CEO Patrick Spence explained: Google and Amazon have flooded the market with dramatically price-subsidized products. Indeed, they make no pretense of the fact that the products themselves are money losers and they routinely give them away at steep discounts, even for free. It is difficult to predict the impact that voice assistants will have on search and e-commerce, but voice activated speakers have the potential to dramatically alter the way that consumers interact with the internet. We believe that Google and Amazon have been willing to forgo profits in smart speakers for this reason, in addition to their ability to monetize the valuable household data that these products vacuum up. And if voice purchasing and voice search do become the next big thing, they will own the market because their strategy is succeeding. Those two companies now control roughly 85% of the U.S. smart speaker market . . . It's not because their hardware businesses are profitable in and of themselves*"[220].

Colonization of ecosystems by data. This approach to voice assistants illustrates the phenomenon of "data colonization of ecosystems," which is also noted by the European Commission in its Internet of Things Final report sector inquiry. Indeed, it points out that user companies do not have "*consistent and immediate access to relevant data about the use of their smart services and devices on third-party voice assistants and smart device operating systems (...). The lack of automatic data collection prevents third-party service providers from personalizing the user experience in real time and puts them at a competitive disadvantage to the top-tier services offered by the leading general-purpose voice assistant providers*[221]".

---

[219] *Ibid.*
[220] *Ibid.*
[221] European Commission, Final report - Sector inquiry into consumer Internet of Things, 20 January 2022, point 483.

2. Relevant data

Data. Companies that do not have access to data are largely dependent on actors that do; those that are able to cross-reference user data and services available on the voice assistant are the only ones able to compete with other actors. In other words, only structuring actors can compete with other structuring actors, because they are the only ones who have a systemic vision[222] of all the services that make up ecosystems, with data flowing through the system as blood flows through the human body. It is no coincidence that the DMA defines data[223], something that Regulation (EU) 2016/679 failed to do in seeking to create categories of data. Beyond that, the importance lies in the data itself, whether it is personal or not. The ecosystem accommodates a global view of data.

The data collected. Digital platforms, including those for voice assistants, concentrate data for purposes such as improving the operation of voice assistants or for commercial purposes. Three categories in particular can be identified for this purpose, which can be classified either as personal data or as non-personal data. First of all, the data used to validate and train the algorithms, collected from text mining, databases, user profiles, etc., are concerned. Secondly, usage data is collected, including personal data, contextual data and other metadata, with the objective of targeting the user's needs, possibly for advertising purposes. As such, the personal data collected are numerous, including the user's contact information (name, e-mail address, location information, IP address) and data relating to his personality, his profession and even his bank details. Finally, the performance data of the voice assistant is used to improve its operation and promote innovation. These data, like the previous ones, include both personal data, and in particular the users' requests, but also metadata and context data, which reveal for example the success or failure of the request or the number of requests made by the user to listen to the same musical extract. Data about the user is collected more generally, whether in the form of information entered by the user or information about the user's behaviour collected in the course of operating the device or accessing the service.

Data control for better market control. In reality, the difficulty of access to data by third-party user companies results from the structuring actor's control of the technical process: from control of the cloud infrastructure in which the data is stored, to development at scale of the algorithmic systems, to the voice application integration process, to technical support, to the creation of the user account that controls the experience with the user. There is no "semantic interoperability" allowing a certain form of standardization of data formats and models, a vertical redistribution of data to user companies who could thus process their performance and usage data, in

---

[222] V. D. Bosco, "A propos du concept de contrôleur d'accès dans le DMA", *Concurrences*, No.3, 2022, p. 64.

[223] DMA, Article 2, "data": "*any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording*".

connection with their own database and using their own tools[224]. It should therefore be promoted.

While access to data allows the voice assistant provider to deploy, the lack of access to that same data creates barriers to new market entrants and competitor development. Respondents to the Commission's sector inquiry on consumer connected objects point out to this effect that voice assistant providers "*have full access to consumer data, while third-party developers have to deal with a really complicated setup process before the user can give us their data, such as their location, to receive more relevant information*[225]". Beyond that, leading providers may be able to degrade the use of services and applications offered by third-party user companies by not fulfilling user requests. At the same time, the voice assistants providers are asking third-party user companies, and more specifically application developers, to transmit a significant amount of metadata to them, particularly as part of the integration procedures on the voice assistant platform. However, application developers have no visibility into how the data is used by the leading voice assistants providers and the outcome of the user's requests. To improve the quality of their service, they need access to both performance and usage data.

Access to usage data. First, the voice assistant provider should not be able to capture all this data without letting third-party application providers have it to manage the experience with the user. In this respect, access to usage data is a key issue. A first step in this direction was taken with the adoption of Regulation 2019/1150. Article 9 of this text in fact falls far short of recognizing a right of access to such data, since it merely establishes an obligation of transparency on the part of platform intermediaries in the form of an obligation to provide information - to be included in their GTCs - on the option they are free to offer their professional users regarding access to such usage data. The DMA goes much further and explicitly enshrines this access to usage data in favour of third-party companies. For example, Article 6.10 requires that *" The gatekeeper shall provide business users and third parties authorised by a business user, at their request, free of charge, with effective, high-quality, continuous and real-time access to, and use of, aggregated and non-aggregated data, including personal data, that is provided for or generated in the context of the use of the relevant core platform services or services provided together with, or in support of, the relevant core platform services by those business users and the end users engaging with the products or services provided by those business users."* With respect to personal data, it states that " *the gatekeeper shall provide for such access to, and use of, personal data only where the data are directly connected with the use effectuated by the end users in respect of the products or services offered by the relevant business user through the relevant core platform service, and when the end users opt in to such sharing by giving their consent.*" The terms and conditions of access to the data will then be decisive. To this end, Recital 60 states that "*gatekeepers should [...] ensure*

---

[224] PEReN, *Shedding light on…N°1 // Interoperability*, October 2021, p. 4.
[225] European Commission, Final report - Sector inquiry supra, point 483.

the continuous and real time access to such data by means of appropriate technical measures, for example by putting in place high quality application programming interfaces or integrated tools for small volume business users".

Access to performance data? With respect to performance data, Article 6.8 of the DMA provides that " *The gatekeeper shall provide advertisers and publishers, as well as third parties authorised by advertisers and publishers, upon their request and free of charge, with access to the performance measuring tools of the gatekeeper and the data necessary for advertisers and publishers to carry out their own independent verification of the advertisements inventory, including aggregated and non-aggregated data. Such data shall be provided in a manner that enables advertisers and publishers to run their own verification and measurement tools to assess the performance of the core platform services provided for by the gatekeepers*". This provision covers the supply of online advertising services. Nevertheless, ensuring contestability and fairness in digital markets following the DMA objective could lead to considering the need to more broadly enshrine a right of access to performance data for the benefit of third-party application providers in order to allow them, when their application is accessible via a voice assistant, to improve the quality of the answer to the question asked by the user and consequently the relevance of their application's ranking. This would also make it possible to evaluate the cases where the user's request has been misunderstood by the voice assistant and in which it would then play, by default, the application of the company which is at the same time the provider of the interface.

Renew a direct relationship with the user. Beyond that, the challenge is to enable application providers to re-establish a more direct relationship with the user who subscribes to their services, which could be another way to access this usage data. This is the meaning of Article 5.5 of the DMA, which states that *" the gatekeeper shall allow end users to access and use, through its core platform services, content, subscriptions, features or other items, by using the software application of a business user, including where those end users acquired such items from the relevant business user without using the core platform services of the gatekeeper."*

3. Data portability

The issue of data portability is often addressed from the user's perspective (see below), but there is an equally interesting aspect in the existing relationship between the voice assistant and third-party user companies with a view to migrating the data centralized by the voice assistant to them. Ideally, voice assistant providers should provide third-party user companies with all data collected in connection with their service on an ongoing basis and free of charge. For the time being, they are using the data in question to develop competing services.

Data portability must be organized within the framework of standardized formats. Otherwise, PEReN specifies that it could be achieved "*using connectors accessible through APIs*"[226]. This approach is already developed between leading actors (see for example the *Data Transfer project*) and to interconnect certain elements of online services (IFTT)[227]. Therefore, the portability of the user's data from one platform to another is possible, and this is done automatically when the user registers, which reduces migration costs. The objective is to allow the decentralization of data, by requiring structuring actors to provide "*asymmetric portability*", which would be obliged to guarantee the various aspects related to data security, such as user consent, confidentiality of personal data and security rules related to consumer protection. These initiatives are still reserved for data migration between specific actors only, but do not yet seem to be targeted at user companies, probably because of the migration costs[228]. To the extent that data portability serves an interoperability purpose, compliance with this obligation may be essential under an *ex ante* regulatory approach.

In the context, data portability to user companies presents an additional challenge in that the aggregation may involve both personal and non-personal data. Therefore, they do not follow the same rules. The user must consent to the portability of his or her personal data to the third-party user company. The conditions for obtaining consent are yet to be specified, as it may be required via the voice assistant interface, or by the user companies[229].

By prohibiting self-preference in order not to block access to the market, by enshrining interoperability in order to promote it, or by guaranteeing access to data in order to open up the market, the reforms of European Union law could support the development of actors in the cultural sector who now use voice assistants to disseminate their offers. At the same time, this action favours access to content, a guarantee of diversity.

---

[226] PEReN, *Shedding light on…N°1 // Interoperability*, October 2021, p.2.
[227] *Ibid.*
[228] *Ibid.*
[229] J. Krämer, *Data access provisions in the DMA*, CERRE, Issue paper, November 2022, p.12.

III - VOICE ASSISTANTS AND ACCESS TO CULTURAL CONTENT

These new forms of human-machine interaction have important consequences on the diffusion and diversity of cultural content. This implies thinking about the instruments to guarantee diversity within these ecosystems.

In order for the user to exercise freedom of choice, it is necessary to question the practices amplified by the use of voice assistants or specific to them in that they alter the user's access to content, particularly with regard to the mechanism of the single answer which will only present him with a single occurrence. The study carried out by the CSA and the HADOPI on connected speakers[230] already noted in this respect a risk of confinement while recognizing the renewed interest of the user for cultural platforms and audio media allowing a better exposure of content. ARCEP's report on terminal neutrality also emphasized this point, stressing the need to guarantee the user's freedom of choice in order to enable him to make informed choices. In effect, voice assistants offer is " *creating an intelligent but internet-centric experience*"[231], with the risk of seeing users "confined" in branded universes distributing terminals, content and possibly services. The operation of these devices also relies on a significant collection of personal data, through the mere recording of voice requests as well as through the analysis of the content of these requests. Thus, the freedom of choice of the user will suppose that the question of consent is adapted to the voice interface, for improved information but also to avoid a "*monetization of intimacy*"[232]. Similarly, as part of its sector inquiry into consumer connected objects[233], the European Commission has outlined its concerns about practices that could impact competition, innovation and user choice. While specifying that at the end of the chain, it concerns the number of voice applications made available to the user, determined by issues of interoperability, default installation of these applications, parametrisation, exclusivity and disintermediation. As has been pointed out, the intuitive nature of user interfaces masks many parameters that are decisive in the computation of query results, which are often limited to a single result whose choice is never neutral[234].

This implies acting on different levers, according to a three-step approach. At first glance, it seems necessary to remove the restrictions on the market by questioning the self-preference practices of voice assistant operators who favour their own services, for example when the voice assistant responds to the user's voice request

---

[230] HADOPI / CSA, *Assistants vocaux et enceintes connectées*, May 2019.
[231] ARCEP, *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open internet*: *Report on their limitations and proposals for corrective measures*, February 2018, p. 31 - https://en.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf
[232] CNIL, *A votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*, White Paper No.1, September 2020.
[233] European Commission, *Final report - Sector inquiry into consumer Internet of Things*, 20 January 2022; see also House of Representatives, *Investigation of competition in Digital Markets*, majority staff reports and recommendations, Part I, July 2022, p. 101 and following.
[234] A. Masure, remarks collected by H. Guillaud, Vox Machines, "Si les assistants vocaux sont la solution, quel est le problème?", *InternetActu*, 19 December 2018.

by favouring its related services (A). Beyond guaranteed access to a plurality of content, freedom of choice also implies ensuring the diversity of this content. This includes considering the technical integration of existing applications to ensure diversity. This makes it possible not only to guarantee access to several contents, which pluralism includes, but also to guarantee access to different contents (B). This supply-side approach to diversity must be complemented by a demand-side approach to diversity. In order to do this, the role of the user needs to be thought of differently, by giving him the means to be informed but also to act on access to content, in particular by giving him the right to set parameters in order to widen his choice of content in the ecosystem, but also to choose this ecosystem by removing the barriers to exit, which presupposes facilitating the conditions for unsubscribing from the service and guaranteeing him a right to portability, and therefore interoperability (C). This means thinking about the mechanisms of co-regulation of diversity by supply and demand.

A - Combating self-preference through the expression of pluralism: the first lever to guarantee freedom of choice

Self-preference, which is in principle a matter of the relationship between the economic actors in the market, is considered here from the user's perspective. The self-preference mechanisms used by the structuring actors have an impact both on the number of voice applications made available, and on the origin of these voice applications, which in essence restricts the user's freedom of choice. There is no need to lock down the ecosystem for the issue of freedom of choice to be analysed from this angle; it is simply a matter of the structuring actor confining the user around its products and services and offering them as a priority at conditions that defy competition.

With the advent of the DMA, the question of content diversity is raised in terms of contestability and fairness of markets. The text establishes one of the facets of transparency and aims to remedy informational asymmetries.

In order to promote their services, business users have to comply with the integration rules set by gatekeepers and their core platform services. A few application stores dominate the market while there are many innovative application stores in the field. In the same way, the offers proposed by voice applications, especially in the cultural area, can be determined by the purchase of a consumer good following a data combination mechanism between the different essential platform services, especially those of the voice assistant. Addressing the question of diversity by starting from the contestability of the construction of the single answer is probably the starting point of the examination. But it only produces a quantitative result, based on the multiplicity of the contents made available, and not on what could be conceived as diversity.

Diversity as a component of transparency. Transparency is a requirement that transcends the most traditional divisions of the legal and scientific communities. This principle must be constantly rebuilt to absorb new challenges and innovation. However, links must now be made between transparency and the implementation of diversity through incentive mechanisms. This incentive may have been direct in terms of applying the principle of pluralism to older models of communication[235]. It is probably built in a more indirect way through information that can compare the offers of services and contents since, from now on, transparency also aims at restoring the freedom of choice of the user through empowerment, in order to preserve his individual liberties as well as democratic debate.

Issues of pluralism. It is necessary to reflect on the way these mechanisms have been exploited in the last century in order to consider how these mechanisms can be adapted to the environment of voice assistants to guarantee the user's freedom of choice. This link between pluralism and freedom of choice was established in its time by the Constitutional Council, to show that without a plurality of actors and contents, the user could not claim to exercise his freedom of choice between diversified contents.

Freedom of choice, diversity and pluralism. For the user to be able to exercise his freedom of choice, it is necessary that he has access to diversified content. This diversity must then be understood from the point of view of pluralism, established first by constitutional law as one of the foundations of democracy, and then by media law, to reflect the different currents of thought and social-cultural opinion. According to the jurisprudence of the Constitutional Council in this area, pluralism must guarantee both diversity or at least plurality of actors (external pluralism), but also diversity or plurality of content (internal pluralism).

Indeed, the Constitutional Council and the ECHR make pluralism a condition of democracy[236], based on the confrontation of ideas and opinions. Pluralism stems from the primary role played by freedom of communication in the functioning of democracy. Moreover, democracy cannot be limited to its political conception. It includes the cultural, religious and social debate without which the appreciation of pluralism would be incomplete. The pluralism referred to differs somewhat, however, when reading the decisions of the Constitutional Council and the ECHR, without being in conflict. Pluralism can mean plurality or diversity of opinions[237]. This difference in perception has an impact on the expected diversity.

---

[235] As regards the press and audiovisual media, see P. Marcangelo-Leos, *Pluralisme et audiovisuel*, LGDJ, Bibliothèque de droit public, t. 240, 2004.
[236] Cons. const., Sept. 18, 1986, no. 86-217 DC; ECHR, Dec. 7, 1976, case 5493/72, Handyside v. United Kingdom.
[237] P. Marcangelo-Leos, supra and "The place of pluralism in audiovisual communication law - The distinction between pluralism and plurality", *Légipresse* June 2003, No.202.

For the Constitutional Council, pluralism is linked to an order of magnitude which, in the area of audiovisual communication and the press, is realised by the preservation of external pluralism (a sufficient number of economic actors to guarantee diversity). The pluralism of contents, qualified as internal pluralism, has only been guaranteed in the audiovisual communication. In the area of the press, the preservation of pluralism presupposes that newspapers are able to publish debates of general interest[238]: the aid distributed by the State is intended to preserve press pluralism by identifying publications with low advertising resources, without really defining diversity[239].

Since the constitutional revision of 23 July 2008, Article 4 of the Constitution has proclaimed "*pluralist expressions of opinion*", but the Constitutional Council limits the application of this principle to political pluralism[240], qualifying other areas as an objective of constitutional value[241]. However, Article 34 of the Constitution was also revised on the same date, and requires the legislator to guarantee the independence and pluralism of the media. The guarantee of pluralism is, at least in appearance, strengthened because it is linked to the independence of the media. In reality, respect for pluralism can be accommodated by its own technical devices, even if they are imperfect, but independence cannot be decreed. Moreover, in an ecosystem, dependence is the primary characteristic of the relationships between actors, which means that the question of independence is built within the framework of an *ex ante* regulatory approach, initially with classic criteria linked to the critical size of the company and the number of users concerned, but from which it would be possible to seek a more refined approach by "patching in" new rules as and when necessary[242]. This is how the DMA (which establishes a form of external pluralism) and the DSA (which establishes a form of internal pluralism) set out their obligations with respect to the largest platforms, but without going so far as to distinguish between obligations according to the services and ecosystems concerned.

Finally, on this principle of pluralism, the Constitutional Council has interpreted Article 11 of the French Constitution, by enshrining the freedom to receive ideas and opinions. In the light of this text, it protects the receiver, imposing on the transmitters the preservation of the pluralist character of the currents of thought and opinion under the control of the CSA as regards audiovisual communication and within the framework of the repressive regime of the press. So that the receivers are "*able to exercise their free choice without private interests or public authorities being able to substitute their own decisions or to make them the object of a market*"[243]. This recital of the Constitutional Council usefully gives value to the receiver's freedom of choice within

---

[238] B. Lamy (de), "La Constitution et la liberté de la presse", *Les Nouveaux Cahiers du Conseil constitutionnel*, No. 36, June 2012.
[239] Or by referring to offences committed via the press to demonstrate that they would impede pluralism. On the basis of social value, a conviction in the last five years for incitement to violence and racial hatred is a bar to receiving assistance: CE, 22 Feb. 2017, req. No. 395948, Société Valmonde.
[240] Cons. const., Jan. 12, 2012, no. 2011-4538 DC.
[241] Cons. const., 28 May 2010, No. 2010-3 QPC.
[242] V. Weigel, "Pluralisme, anti-concentration et média global", *RLDI* 2008, p. 44.
[243] Cons. const., Oct. 10 and 11, 1984, No. 84-181 DC, Media company; Cons. const., Sept. 18, 1986, No. 86-217 DC, Freedom of audiovisual communication.

the framework of what the transmitter proposes to him but not within the framework of what he intends to choose. This freedom of choice was already conditional in the twentieth century; it is now more so, giving the receiver the illusion of choice because of the question of access to services and their visibility.

Pluralism in the 21st century. The principle of pluralism as it was constructed by the Constitutional Council is frozen in contemplation of the linear practices of the 20th century, but does not prevent us from reflecting on the adaptation of this principle to digital uses, and particularly to voice assistants. Pluralism must therefore guarantee a plurality of actors in the market (1), a diversity of ways of accessing content (2) and pluralism of content (3).

1. Plurality of actors on the market

Two significant examples are discussed here. The first puts into perspective the ineffectiveness of concentration thresholds in guaranteeing external pluralism in the face of structuring actors that fall below the thresholds. The second one questions respect for external pluralism because of the reappraisal of radio signal reception by the new generation of cars.

Plurality of actors and external pluralism. The example of compliance with concentration limits. The control of company concentration thresholds, a mechanism used by both EU [244] and national law [245] to guarantee respect for external pluralism,

---

[244] Article 1 and recital 20 of the Regulation No.139/2004 of 20 January 2004 on the control of concentrations.
Recital 20: (20) " *It is expedient to define the concept of concentration in such a manner as to cover operations bringing about a lasting change in the control of the undertakings concerned and therefore in the structure of the market. It is therefore appropriate to include, within the scope of this Regulation, all joint ventures performing on a lasting basis all the functions of an autonomous economic entity. It is moreover appropriate to treat as a single concentration transactions that are closely connected in that they are linked by condition or take the form of a series of transactions in securities taking place within a reasonably short period of time".*
Article 1: "*Scope: 1. Without prejudice to Article 4(5) and Article 22, this Regulation shall apply to all concentrations with a Community dimension as defined in this Article.*
1. Without prejudice to Article 4(5) and Article 22, this Regulation shall apply to all concentrations with a Community dimension as defined in this Article.
2. A concentration has a Community dimension where:
(a) the combined aggregate worldwide turnover of all the undertakings concerned is more than EUR 5000 million; and
(b) the aggregate Community-wide turnover of each of at least two of the undertakings concerned is more than EUR 250 million, unless each of the undertakings concerned achieves more than two-thirds of its aggregate Community-wide turnover within one and the same Member State.
3. A concentration that does not meet the thresholds laid down in paragraph 2 has a Community dimension where:
(a) the combined aggregate worldwide turnover of all the undertakings concerned is more than EUR 2500 million;
(b) in each of at least three Member States, the combined aggregate turnover of all the undertakings concerned is more than EUR 100 million;
(c) in each of at least three Member States included for the purpose of point (b), the aggregate turnover of each of at least two of the undertakings concerned is more than EUR 25 million; and
(d) the aggregate Community-wide turnover of each of at least two of the undertakings concerned is more than EUR 100 million, unless each of the undertakings concerned achieves more than two-thirds of its aggregate Community-wide turnover within one and the same Member State. 4. On the basis of statistical data that may be regularly provided by the Member States, the Commission shall report to the Council on the operation of the thresholds and criteria set out in paragraphs 2 and 3 by 1 July 2009 and may present proposals pursuant to paragraph 5. 5. Following the report referred to in paragraph 4 and on a proposal from the Commission, the Council, acting by a qualified majority, may revise the thresholds and criteria mentioned in paragraph 3".
[245] Article L. 430-2 of the Commercial Code: "I.- Any merger, within the meaning of Article L. 430-1 is subject to the provisions of Articles L.430-3 et seq. of this Title when the following three conditions are met:

raises questions about the effectiveness of these thresholds in revealing the market position of a digital player when turnover is not related to the valuation of the target. This question was already present in the 20th century to evaluate the effects of controlling concentration thresholds on internal pluralism (of content). The capital thresholds set by the law of 30 September 1986, as amended, could not be assessed as such, without taking into account the audience criteria for analysing the editorial impact of an audiovisual communication company on the characterization of content pluralism[246]. The case of Apple's acquisition of the Shazam application illustrates the ineffectiveness of such thresholds when a company that does not fall within the scope of the 2004 regulation is nevertheless valued at $1 billion. For this reason, the Austrian Competition Authority, joined by six national competition authorities including France, requested a referral to the European Commission under Article 22 of the regulation to re-examine the situation of the takeover, a sign of the European Union's reaction to the increasing number of acquisitions of innovative companies with high value but low turnover.[247] Taking into account the information submitted by Austria and the States joining the referral request, and without prejudice to the outcome of its in-depth investigation, the Commission has considered that the transaction may have a significant adverse effect on competition in the European Economic Area. The Commission also concluded that it was the best placed authority to deal with the potential cross-border effects of the transaction. The Commission has asked Apple to notify it of the transaction. For Philippe Corruble, *"this phenomenon appeared all the more painful as the acquisition was made, in certain transactions, by a powerful or even dominant buyer and for a very high amount. Thus, transactions as important as Facebook/Instagram, Google/Looker, Amazon/Whole Foods or Google/Waze escaped any control by the European Commission: the regulation simply did not apply. This situation clearly called for an awareness of the need for a change in approach (Cousin, 2019). It was first addressed by some national competition authorities"*[248]. Article 14 of the DMA will require gatekeepers to notify the Commission of any proposed merger, where the merging entities or the target of the acquisition provide core platform services or any other digital service, or allow the collection of data. This obligation will be imposed from 2023 onwards regardless of whether the transaction is to be notified to the Commission under the merger regulation or to a national competition authority under national rules, which is reminiscent of the procedure under

---

-the total worldwide turnover excluding tax of all the undertakings or groups of natural persons or legal entities involved in the merger is greater than 150 million euros; - the total turnover excluding tax earned in France by at least two of the undertakings or groups of natural persons or legal entities concerned is greater than 50 million euros; - the transaction does not fall within the scope of Council Regulation (EC) No. 139/2004 of 20 January 2004 on the control of mergers between undertakings. II - When at least two of the parties to the merge operate one or more retail stores, the provisions of Articles L. 430-3 et seq. of this title, and any merger within the meaning of Article L. 430-1, when the following three conditions are met: the total worldwide turnover (excluding tax) of all the undertakings or groups of natural or legal persons involved in the merger exceeds EUR 75 million; - the total turnover (excluding tax) in France in the retail sector of at least two of the undertakings or groups of natural or legal persons involved exceeds EUR 15 million; - the transaction does not fall within the scope of the abovementioned Council Regulation (EC) No. 139/2004 of 20 January 2004.

[246] K. Favro, *Téléspectateur et message audiovisuel - Contribution à l'étude des droits du téléspectateur*, LGDJ, Bibliothèque de droit public, volume 215, 2001, p.89 and following.

[247] Commission Decision, 6 Sep. 2018 Apple/Shazam. M.8788; European Commission Press Release No. IP/18/664, 06 February 2018.

[248] P. Corruble, "Le contrôle européen des concentrations à l'épreuve de l'Economie numérique", *Revue Interdisciplinaire Droit et Organisation*, Association RIDO, 2022, 3, p.25, 10.34699/rido.2022.11. hal-03736149.

article 22 of the merger regulation but with more appropriate criteria than those resulting from the 2004 Regulation. The thresholds imposed by competition law do not guarantee external pluralism. A regulatory approach that puts qualitative rather than quantitative criteria into perspective is expected through the application of the DMA (see above).

External pluralism and radio signal. Voice assistants and connected speakers must not deprive viewers and listeners of receiving a radio signal, especially when that is their choice. The problem may be transitory, but linear broadcasting still makes up 20% of the market, and the disappearance of the radio frequency would be an attack on external pluralism in that it would deprive listeners of their freedom of choice. Nevertheless, this issue of a "*technological mix*" for access to linear radio and television services is currently being played out, especially for radio, as it depends largely on vehicle equipment. If car manufacturers decide to stop equipping vehicles with antennas to receive radio signal, not only will listeners lose their freedom of choice and free access to radio services, which will be conditioned by the cost of a subscription to a network or even a subscription to services, they will also be equipped with a dashboard that will engender other difficulties, in particular accessing their preferred services. Maintaining linear broadcasting frequencies for radio and television will help maintain the universality of these free services and anonymity in choice for audiences.

2. Variety of ways to access content

Maintaining diversity in the ways users can access content implies that attention should be paid to what is at stake in single responses and access to equipment. More specifically, it requires us to consider the interfaces that are included in new generation vehicles.

Pluralism and single responses. Single responses provided by voice assistants should encourage the search for tools that would allow upstream of the request to classify content, because of their general interest, in order to guarantee democratic debate. The long-term objective would be to propose specific obligations either for voice assistant creators or for services provided by user companies.

End-users cannot specify the service they wish to use as part of their requests. The virtual assistant will then suggest a default service. For instance, for the news updates, available public service content (Radio France) is currently launched by default. On Google Assistant and Amazon Echo, radio services are launched by default through the TuneIn aggregator. It is up to users to change this selection of voice applications in the settings if they wish to access other content. In practice, users rarely install applications that offer the same functionality as an application installed by default,

unless they consider that it does not have the expected quality, or that it involves an unwanted subscription. Therefore, the default installation can be analyzed as meeting the expectations of users who would like to start using their voice assistant out of the box. Moreover, the answer given will always depend on how a question is formulated, rankings made by algorithms, software building blocks that put intentions in the answer, and the framework of the answer (is it simply vocal?).

Virtual assistant manufacturers do not require any exclusivity from content publishers, so theoretically they can be present in all ecosystems. But some partnerships are difficult to circumvent "*because the limited nature of the content choice interface on the speakers makes the choice of manufacturer even more decisive*"[249]. In addition, the development costs of the voice applications built into each assistant's system are significant for third-party companies. Furthermore, the system remains opaque in that core platform services have the ability to give preference to their own services by cross-referencing data between core platform services, and in particular marketplaces or app stores, to give them a competitive advantage over third-party companies.

In a way, remedying self-preference through the application of the DMA amounts to quantitatively guaranteeing the user's freedom of choice. This is what should be retained from Article 6.3 of the text: "*The gatekeeper shall allow and technically enable end users to easily un-install any software applications on the operating system of the gatekeeper, without prejudice to the possibility for that gatekeeper to restrict such un-installation in relation to software applications that are essential for the functioning of the operating system or of the device and which cannot technically be offered on a standalone basis by third parties. The gatekeeper shall allow and technically enable end users to easily change default settings on the operating system, virtual assistant and web browser of the gatekeeper that direct or steer end users to products or services provided by the gatekeeper. That includes prompting end users, at the moment of the end users' first use of an online search engine, virtual assistant or web browser of the gatekeeper listed in the designation decision pursuant to Article 3(9), to choose, from a list of the main available service providers, the online search engine, virtual assistant or web browser to which the operating system of the gatekeeper directs or steers users by default, and the online search engine to which the virtual assistant and the web browser of the gatekeeper directs or steers users by default*".

This article sets out the preconditions for taking the approach further and targeting the services offered and fighting against proprietary ecosystems. It suggests that a voice assistant could be delivered "blank" or possibly with the ability to disable or "reconfigure" the voice applications installed by default. Article 6.3 also allows users to make online purchases after selecting an operating system and voice assistant of their choice. It does so by using the voice application of their choice and not orienting them to specific platforms (marketplaces and application stores) related to the voice assistant ecosystem through AI content rankings.

---

[249] ARCEP, *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open internet: Report on their limitations and proposals for corrective measures*, February 2018, p.42.

Finally, Recitals 62 and 63, which clarify the meaning of this article, focus on the ability of users to unsubscribe, which makes it possible to make the preventive mechanisms for self-referral effective. Indeed, it states, "*To safeguard free choice of business users and end users, a gatekeeper should not be allowed to make it unnecessarily difficult or complicated for business users or end users to unsubscribe from a core platform service. Closing an account or un-subscribing should not be made be more complicated than opening an account or subscribing to the same service*".

As a mechanism to remove systems and techniques that discourage competition, the simple application of the DMA to voice assistants would make it possible to guarantee users' freedom of choice with respect to the default voice applications that they have not uninstalled and those that they have installed, which, without any pre-established ranking, will be made available to them via voice command. However, the obligations set out to establish free competition only police the system in a preventive manner without any real asymmetry, apart from Article 8 of the DMA, which provides for the partial application of obligations depending on the issues at stake. Their primary purpose is not to achieve diversity. Only a genuine approach based on asymmetrical ex ante regulation would make it possible to impose obligations - obligations to be adjusted based on the diversity objective - on all gatekeepers with respect to their core platform services.

But this freedom of choice remains conditional: it depends on technical building blocks, on disinformation, and on devices that struggle to guarantee cultural pluralism.

Freedom of choice and access to devices. The European Open Internet Regulation[250] considers the Internet to be "*an open innovation platform that is easily accessible to end users, content, application, and service providers, and Internet access service providers*". The Constitutional Council, in the context of the HADOPI[251] law, considered, on the basis of Article 11 of the French Constitution, that participation in democratic life and the expression of ideas and opinions "*implies the freedom to access these services*" On this basis, the law grants rights to users, namely the right to access and distribute information and content online on the device of their choice. However, the open Internet is only considered from the point of view of the neutrality of Internet service provider networks (traffic management practices, transparency) and only aims to prohibit restrictions on the use of devices connected to networks.

However, starting in February 2018 and following an inquiry conducted between 2015 and 2018, ARCEP showed that devices play a significant role in the market by offering content. A follow-up to that finding is reflected in the provisions of the DMA. This is the case for voice assistants, to which ARCEP was clearly referring, as well as for

---

[250] Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures on access to an open internet.
[251] Decision No. 2009-580 DC of 10 June 2009 is a law to promote the dissemination and protection of creative work on the Internet.

operating systems, browsers, and application stores. Access to the user's network with the help of the voice assistant is indeed conditioned by the existence of many software building blocks. For industry actors, what is at stake here is data exchange. The first consequence is that the user is not able to appreciate all the functionalities of his voice assistant out of the box. Therefore, the user may naturally adopt manufacturer choices without being able to analyse them. Naturally, however, the lack of a visual display due to voice interaction limits the possibilities of accessing comprehensive information and leads users to select information presented by user companies since voice applications have to be developed. These limits result from the nature of the object but distort access to the Internet by fragmenting it. It is worth noting that voice assistant interface limitations are reflected on all the devices to which it provides access. This is particularly the case when it comes to smart speakers insofar as not all the content they could broadcast is intended to pass through the speaker in the absence of a screen. We can consider that these speakers give access to all consumer connected objects by serving as a relay. Or conversely, we can say that the voice speaker limits integration to voice applications within the framework of a simple transparent interfacing system known by the users (e.g. Sonos).

The challenge of new generation vehicles. The development of voice applications may also accompany new uses. The automotive industry is a good example: manufacturers are now integrating "*voice-first technology into their new models for technical commands related to the operation of the vehicle*" and "*embedding voice assistants to manage entertainment (music, radio, etc.) and organizational functions (calls, appointments, etc.), either by developing them internally or by signing partnerships with technology companies, such as Amazon (Echo Auto), Apple (Car Play) and Google (Android Auto)*"[252]. While the car has always been an important market for radio broadcasters, they now have to position themselves towards manufacturers to appear on the dashboards of the latest generation of vehicles. The challenge is essential as new generation vehicles are equipped with dashboards that allow voice access to various services and content, particularly radio services and music subscription services. Since the majority of radio use is mobile, radio service providers must be visible and referenced on the vehicle dashboard. This requires them to carry out major sales negotiations, particularly with vehicle manufacturers, in order to appear on the dashboard. This question is now being asked of manufacturers of connected TVs regarding the placement of certain services on the remote control "buttons" of these TVs. A battle has started over the positioning of audiovisual service publishers on these interfaces. It is then the manufacturers of these connected objects who play the role of intermediaries between these services and the user. There are two issues for service providers: the plurality of negotiations to be conducted when there are a large number of manufacturers of connected objects; the difficulty of conducting these negotiations in a balanced way, especially since these manufacturers operate globally and may not want to adapt each of these interfaces

---

[252] O. Gouliaéva, E. Dosquet, Y. Moysan, *La révolution des assistants vocaux*, DUNOD, 2020, p. 59.

every given market. With regard specifically to vehicle dashboards replacing traditional car radios, one idea would be to impose wireless access being upheld to guarantee access to the service that does not appear on the dashboard "button". Another solution would be to allow the user to configure the dashboard from a list of services (see *below*). Informed users are not unaware of these issues, but naturally turn to equipment that meets their use habits online. The fact is that this choice, even in the absence of deliberate action by professionals, locks users into an ecosystem and into a habit of fragmented listening. However, the choice of voice assistant must be dissociated from the choice of content and services used. The objective is that in the long run they become equivalent in application of the principle of non-discrimination.

3. Plurality of content

From honest information to misinformation. The principle of internal pluralism, which implies diversity or at least plurality of content, can only be understood in the context of providing honest information. In the field of audiovisual communication, the principle of honesty of information is part and parcel with the preservation of pluralism. Guaranteed in the field of audiovisual communication, it has the value of a constitutional imperative[253], which the CSA and now the ARCOM has the mission to monitor in the same way as for pluralism[254]. The imperative, which is strongly ethical in nature, cannot be derogated from in any way and is opposed to requirements that would only be random in scope.[255] Generally speaking, the agreements drawn up by the CSA and now by ARCOM are very specific on this subject and impose the respect of this principle for all programs.[256] Honesty is not a principle of truth, but it presumes that someone who makes information available respects the truth both in content and in form.[257] Honesty also implies gathering different points of view even if they are contradictory, the objective being that the public can form its opinion and make choices in an informed way. This objective has disappeared from the terminology used by the legislator, who now prefers to speak of issues of disinformation. However, the scope of misinformation remains the same, targeting both the content and the way it is made available to the user, which may raise questions about the visibility of interfaces and illicit comments or the issue of quotas. Moreover, processes of disinformation are sometimes intertwined. A comprehensive perspective, with honesty on the positive side of the coin and misinformation on the negative side, can provide insight into all the facets of user freedom of choice and encompass contemporary issues through a choice of terminology.

---

[253] Cons. const., No. 86-217 DC, supra.
[254] Combined reading of articles 1 and 28 of the aforementioned law No. 86-1067 of September 30, 1986, supra.
[255] V. A. Guedj, *Liberté et responsabilité du journaliste dans l'ordre juridique européen et international*, Univ. Paris II, 2000.
[256] G. Weigel, "Les dispositions audiovisuelles de la loi 'indépendance des médias' du 14 novembre 2016 étaient-elles nécessaires ?", supra.
[257] See the case of the "fake interview with Fidel Castro", CA Paris, 1st ch. A, 24 Feb. 1998.

Freedom of choice and misinformation. The technical limitations of the voice assistant, which are destined to disappear in the context of a more elaborate virtual interface, may contribute to misinforming users. Moreover, the production of answers and their restitution via a synthetic voice are the result of a long and opaque process that can be understood from the point of view of deceptive choice interfaces (dark patterns), targeted advertising, and other illicit process likely to mislead users in their ability to make choices. The rise of misinformation is another phenomenon linked to large platforms, which moderate information by downgrading content according to systems that are not clear to users.

The fight against disinformation should be based on another, more qualitative approach of pluralism, on which the ECHR relies. In European law, the notion of democracy, although it represents a fundamental aspect of the European public order, has not received a precise definition, given the diversity of national cultures. The Court has developed a jurisprudence that is "*more insistent on the rights of minorities than on the power of the majority*"[258] and contributes to making all ideas heard, even those that offend and those that shock, insofar as pluralism expresses tolerance and open-mindedness[259]. This approach is interesting in that it describes the issue of discoverability of content and reflects on its legality, because pluralism cannot be understood in terms of illicit content.

Under the DSA, illegal content "*cover information relating to illegal content, products, services and activities. In particular, that concept should be understood to refer to information, irrespective of its form, that under the applicable law is either itself illegal, such as illegal hate speech or terrorist content and unlawful discriminatory content, or that the applicable rules render illegal in view of the fact that it relates to illegal activities. Illustrative examples include the sharing of images depicting child sexual abuse, the unlawful non-consensual sharing of private images, online stalking, the sale of non-compliant or counterfeit products, the sale of products or the provision of services in infringement of consumer protection law, the non-authorised use of copyright protected material, the illegal offer of accommodation services or the illegal sale of live animals. In contrast, an eyewitness video of a potential crime should not be considered to constitute illegal content, merely because it depicts an illegal act, where recording or disseminating such a video to the public is not illegal under national or Union law. In this regard, it is immaterial whether the illegality of the information or activity results from Union law or from national law that is in compliance with Union law and what the precise nature or subject matter is of the law in question*"[260]. It is from the point of view of access to safe, predictable, and reliable information that we should fight against disinformation in all its forms, and not from the point of view of content per se. The objective of the DSA is not to police the net and eradicate all disturbing content or fraudulent services, but to limit them and work upstream on behaviours so

---

[258] E. Casenove, *Ordre juridique et démocratie dans la jurisprudence de la CEDH*, Univ. Amiens, 1994, p. 161.
[259] ECHR, 7 Dec. 1976, Handyside v. United Kingdom, supra.
[260] Regulation 2022/2065, Recital 12.

that such content disappears progressively by *ex ante* regulation mechanisms. However, choices must be made according to the content and services targeted, and the behaviour of intermediaries. In this respect, it should be noted that the texts recently adopted or still under discussion, concerning press offenses or even the fight against infringing content, raise converging questions (What is the responsibility of platforms? How should content by flagged? What is the impact of algorithmic filtering? How do we articulate various fundamental rights?) Some of these questions can be answered ex post. However, for the most part, an *ex ante* identification of illicit processes can help distinguish between appropriate responses, methods to use, and the actors involved. This is the difficulty, because as the DSA stands, the European legislator is targeting both platforms and illegal content/services, but without taking a comprehensive view. The difficulty lies in the fact that the dysfunction may not be due solely to the functioning of the platforms or the circulation of illicit content, but to the system as a whole, which must be better understood in order to better manage. Therefore, the DSA was not initially designed to understand the voice assistant or the whole system it is supposed to regulate.

For voice assistants, the issue of disinformation can be understood in terms of the promotion of some types of content whose illicit character can result from both the content (and in particular the promotion of partisan content by the government during the first lockdown) as well as from the fraudulent capture of known voices to boost legitimacy. It can also be understood from the point of view of promoting certain services during the preliminary choices offered when the service is being configured. Responses are therefore multipronged.

Based on ARCEP's recommendations, a first global response could be to regulate through data, because data helps to irrigate the system and it would be advisable for there to be sufficient data to inform all technical intermediaries in order to guarantee this diversity by making available reliable content that respects fundamental rights (see below).

But, more symptomatically, the approach would be to instead promote some services considered as general interest in order to make reliable services and sources visible. In this respect, the revised "Audiovisual Media Services Directive (AVMSD)" proposes[261] rules on the visibility of content qualified as being of general interest. It provides that Member States are free to take measures to ensure the appropriate visibility of audiovisual media services of general interest (Article 7a).

Freedom of choice and access to cultural content. Preserving the user's freedom of choice from the point of view of pluralism could also mean preserving the diversity of the offering in the cultural field, through the respect of quotas. With regard to the

---

[261]Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (AVMSD) in view of changing market realities.

broadcasting of audiovisual content, a combined reading of Articles 33 and 70 of Law 86-1067 of 30 September 1986, as amended, specifies that service publishers should broadcast at least 60% original European works, including 40% original French works. At the European level, the ADM Directive (EU) 2018/1808 also implements quotas. Starting with a first initiative known as *Television without Frontiers* 89/552/EEC of 3 October 1989, the European Economic Community instituted a system of quotas to television and radio channels, through a proposal by France, so as to oblige audiovisual communication companies to protect and establish European heritage by broadcasting at least 50% European works. At the time, it was a linear market that was controlled by the Member States because of the limited number of actors. The objective of the system was to build a true internal market for audiovisual services by breaking down geographical borders. The change in logic brought about in 2007 by the amendment 2007/65/EC of the European Parliament and of the Council of 11 December 2007, in response to shifts in service due to the de-linearization of audiovisual programming, gave rise to AVMS but did not call into question the protective intention of the States. In the 2018 amendment, this arrangement was completed to cover AVMS. On-demand audiovisual media service providers must ensure that their catalogues contain a minimum of 30% European works and that they are sufficiently highlighted by algorithmic systems. The quotas of French music on the radio have undeniably allowed French-speaking artists to have a wide exposure. The question of quotas clearly remains central to the regulations, which are not intended to preserve cultural diversity but simply competition, and in a manner that varies greatly depending on whether the country is a member of the European Union or a third country.

Quotas: an outdated mechanism? Successive amendments to the European system have neither had the effect of sustainably combating forum shopping, nor of preserving a European or French culture within the ecosystems. The CSPLA's mission on the recommendation of works to users of online platforms, led by professors Valérie-Laure Benabou and Joëlle Farchy[262], shows that quotas are not very useful for music, and that they are unsuitable for platform services whenever it is easy to meet demand beyond the regulatory requirements by means of recommendations. This is what Amazon and Netflix are doing, with the help of German and Italian users, who are particularly fond of European content. The single response system, which goes beyond a simple recommendation, reinforces this observation. The report highlights that diversity, as reflected in these regulations, is not adapted to the on-demand service economy. The issue is not of stocking works to meet demand; they have to be discoverable. The presence of content does not mean that it is discoverable, especially considering the single response mechanism. The voice assistant does not inform the user of its reasoning process or the scope of its choices. Nor does it clarify the services activated in the "production" pf a response. The ability to filter content accentuates the risk of not allowing services to emerge within the high-

---

[262] V.-L. Benabou, J. Farchy, *Les dispositifs de recommandation des œuvres audiovisuelles et musicales sur les services en ligne*, report presented to CSPLA, November 2021, p. 64 and following.

volume offering. The negotiating power of voice assistant publishers puts them in a position to capture a significant portion of the value and to impose the conditions under which this value will be returned. There is then a real obstacle to diversity due to SEO that is correlated to the user's formal query. If the user does not enrich the request by making a simple request, diversity will be further impacted.[263] Indeed, a request is said to be closed when the user specifies a specific service in a request. In this context, SEO is always limited in essence. The difficulty then arises from so-called "open" requests, when the service to be launched for a specific action is not specified. The object of the request is then limited to "*Play me this song*" or "*Launch a news update*". SEO is now fundamental because the practice favours the best-known brands and leading providers, to the detriment of emerging actors or those with limited financial capacities, even though they offer more diversified content without being able to justify high listening statistics. This is where the diversity issue lies, apart from any quantitative assessment that helps to establish the content of the structuring actors. A relic of linear television, this quota system does not make sense with regard to the use of voice assistants, for which questions of discoverability are very important.

Can quantitative criteria be replaced by qualitative criteria? Therefore, the avenues proposed in the recommendations report but immediately refuted by the authors must, at the very least, be further studied. Assessing diversity is undoubtedly a matter of political choice. This is the case with the current quota system, which advocates for protectionism. However, this system can only be successful when applied to a circumscribed territory with a landscape of actors controlled by the regulatory authorities. In the absence of quotas, protectionism could be imposed through how algorithms are configured, as well as language diversity, novelty, emerging talent, the presence of minorities and women, or budget diversity, as the authors of the report point out. Criteria like fame (which has long been used as a criterion in cultural programming on television), regional origin (for live shows), and public information in an effort to fight against disinformation are parameters to take into consideration. In this respect, it is not up to the legislator to set up a new system but to the regulator to verify that these editorial constraints have been taken into consideration and in ways that cohere with the services in question. The question is how to reconcile them with personalization and how the regulator will have the means of enforcement while respecting user choice.

Visibility and cultural diversity (Ordinance of 21 December 2020 transposing the AVMS Directive). Formally, it would be a matter of checking that such content is automatically taken into account in the "production" of the response by the voice assistant and in particular during the first "service activation" phase before the voice assistant delivers its response. One idea would be to create services of general interest in line with the model of Article 7 of the AVMS directive. The transposition of the directive by the ordinance of 21 December 2020 entrusts the CSA (ARCOM) with

---

[263] HADOPI / CSA, Study supra, p.65.

new missions regarding the protection of the integrity of the signal of audiovisual communication services and the promotion of audiovisual services of general interest on the new audiovisual content access interfaces. That was the basis for which, in 2020, the European Broadcasting Union (EBU) called on Member States to ensure the visibility of general interest content on audiovisual platforms to "*guarantee the right of citizens to receive accurate and impartial information and to combat misinformation. Without visibility, local content, content aimed at minorities with specialized audiences, and other informative genres, (...) are not promoted by the algorithms. General interest content should extend to disaggregated items of content.*"

Indeed, Article 10 of the Ordinance, which can be found in Article 20-7-I of Law 86-1067 of 30 September 1986, amended, provides that, "*For the purposes of this article, 'user interface' means any device that offers the user a choice among several audiovisual communication services or among programs from these services.* A device installed on a television set or on equipment intended to be connected to the television set; installed on a connected speaker; made available by a service distributor or made available within an application store, does not exclude voice assistants. It is obligated to provide services of general interest under the conditions specified by ARCOM, which must ensure their visibility (except if the interface offers exclusively the services of a publisher and its subsidiaries). These services include those produced by public broadcasting organisations and even other publishers of audiovisual communication services if they are likely to broadcast such content. In practice, a proliferation of publishers of private audiovisual communication services seems desirable in order to guarantee their contribution to the "*pluralistic character of currents of thought and opinion and to cultural diversity*", particularly in areas where only access to public service content is guaranteed.

Under the terms of the ordinance, as of 1 January 2022, operators who determine how services are presented on user interfaces must ensure appropriate visibility of general interest services whose number of users or units marketed on French territory exceeds a threshold set by decree. To our knowledge, no decree has specified the number of users concerned or the number of units marketed. The list of these services is not published. Visibility should be appropriate to the user's ability to customize the interface. Therefore, depending on the case, these services could appear: on the home page or screen, in recommendations, in user-initiated search services, or on *remote control devices for equipment providing access to audiovisual communication services*. The operator who manages the interface must report to ARCOM, under penalty of sanctions after formal notice.

Although interesting, this system is not complete, particularly because of the absence of an implementing decree. Several difficulties appear regarding its implementation: what kind of visibility would be possible for voice assistant interfaces that would allow for the broadcast of these types of content? Which voice applications could be considered to be services of general interest? In fact, the difficulty of listing services

of general interest is probably the crux of the system. For example, out of the 1,000 French radio stations, 700 are associations. Could we not consider all radio associations to be services of general interest because of their purpose and method of funding? In which case, that would mean that all radio services are potentially services of general interest. So, it is difficult to pursue the discussion. And finally, one of the pitfalls of this directive is that it upholds the country of origin rule, when it is likely through an application of the country of destination rule that would guarantee the user's freedom of choice. However, the arrangement should not be discarded since it should be combined with the user's action at the time of configuring the settings (see below) or, more precisely, initiating the service. A conversation between the machine and the human would be necessary to foster diversity. Moreover, it will probably have to be crossed with other data-related measures that could be qualified as being of general interest, as envisaged by the adopted text of the regulation on European data governance[264].

All of the rules presented are based on thresholds and quotas that, because of their objective nature, do not make it possible to grasp the reality, namely the extension of ecosystems through the use of voice assistants, and then the diversity of content through the full application of the requirement for transparency, particularly with regard to the principle of pluralism, which is still measured by an order of magnitude, whose compliance is incumbent on the issuer. Today, this principle seems to integrate the European reasoning on market challenges, honesty of information, and content diversity. Diversity remains to be built by a qualitative appreciation not of the market this time, but of the content/services. The user contributes greatly to making this diversity. This qualitative assessment cannot be based solely on the market's structuring actors. In a regulatory approach advocated by European texts, it is necessarily diluted over all the actors, including users, and is not based solely on one system but on a set of systems. User information and choice should help define diversity.

B - The challenge of discoverability: a second lever to consider diversity differently

What we have just grasped in terms of pluralism through the legibility of interfaces implies continuing the process and rethinking diversity from the point of view of content discoverability. This is a sensitive subject because it reflects a real paradox: the offering of content has never been so abundant on all online services, be it in the form of a revival of the linear television signal, SVOD, or catch-up TV[265]. In the field of music, supply exceeds demand and the market is dominated by a handful of actors.

---

[264] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).
[265] Joint study conducted by the HADOPI and the CSA, HADOPI / CSA, *La multiplication des services de vidéo à la demande par abonnement : stratégies de développement et impact sur les usages*, 28 May 2019, published on 9 March 2021.

Content is therefore available, but is it all accessible to the user? Addressing this question in the world of voice assistants implies highlighting the availability of voice applications for the user, but also going back to the filter of the single answer, which is vocal. The issue of discoverability is then mainly related to the number of voice applications available.

The availability of content for the user does not imply its discoverability. However, work exposure does depend on it[266]. A definition has been proposed in the framework of a French-Canadian Mission: "*Discoverability of content in the digital environment refers to its availability online and its ability to be found among a large body of other content, especially by someone who was not specifically looking for it*"[267]. Apart from important points that can limit the discoverability of content but that are not specific to voice assistants (such as the quality of metadata associated with content, marketing and design strategies, recommendations[268], knowledge of usage by professionals, or public policies promoting such or such types of content), other specific features contribute to locking the user into constrained choices. This is true for the number of voice applications, the vocal character of the response, and the single voice response. The effects are particularly amplified in the unique response to the user. The definition can then be adjusted to apply to voice assistants, with the specification that, "*The discoverability of content in the digital environment refers to its availability online and its capacity to be found among a large set of other content, in particular by a person not formulating a specific request or question*".

The risk of circumscription. The conclusions of Joëlle Farchy and Valérie-Laure Benabou in the context of their report on recommendation mechanisms question[269] the effectiveness of such mechanisms. The proven risk remains that of a system's self-preference linked to content ranking. Narrow circumscription due to the recommendation has not been demonstrated. Assuming that there is an "echo chamber" in the field of access to information, fed by selection biases and self-reinforcing mechanisms of the user's convictions, it only confirms the user's natural tendencies, valuing information that corresponds to the user's convictions. However, the recommendation may have a greater amplifying effect if the response elements are limited to a single answer. For cultural services, the economic interest lies in the immediate satisfaction of user needs, so much so that recommendation algorithms focus on the relevance of results and are not very efficient in the long term, at the risk of boring the user searching for surprises. Boredom is further amplified by the response the user receives after profiling.

---

[266] Report by S. Clément-Cuzin, Inspector General of Cultural Affairs and F. Hurard, Inspector General of Cultural Affairs, *La captation de spectacles vivants*, June 2022, IGAC Report, No. 2022-12.
[267] https://www.culture.gouv.fr/Thematiques/Europe-et-international/Decouvrabilite-en-ligne-des-contenus-culturels-francophones
[268] Idem, p.3.
[269] V.-L. Benabou, J. Farchy, *Les dispositifs de recommandation des œuvres audiovisuelles et musicales sur les services en ligne*, report presented to CSPLA, November 2021, p. 62 and following.

This could lead the user to change applications, or suggest other content by formulating a new request. However, the request must be specific. Because everything will depend on the voice request. The way it is formulated will guide the voice assistant in how it understands the request and identifies available answers or expresses its lack of understanding with respect to the request. Depending on the technology used and the user interface, either the voice assistant identifies the appropriate option (the answer) or a list of options (a list of possible answers). Nevertheless, the European Commission in its consumer IoT sector inquiry has found that to access a non-default service provider via a voice assistant, a user will typically need to add a specific name to their voice commands (e.g. "listen to song X on service Y"). This finding is corroborated by inquiry respondents: "*We have learned that voice command users use very generic commands and seldom include brands in their commands. In our case, 'generic' searches for music channels are directed to the 'preferential' […] app. Only when consumers ask very specific questions for our services, they get directed to us, making 'discoverability' of our services very hard*"[270].

The limited number of voice applications. However, when it comes to the response provided by the voice assistant, the availability of content is as much in question as its visibility. Content availability is linked to how the device is used or to security features limiting access to information through voice activation[271]. In addition, the range of services offered is still insufficient, which constrains the user. The HADOPI-CSA study on connected speakers states that, "*the choice of the brand of the connected speaker has a significant impact on subscriptions to cultural offerings that will subsequently be associated with the device* (....) *In the end, 47% of music subscribers (or 15% of regular users) and 50% of VOD subscribers who use connected speakers (or 21% of regular users) were influenced in their choice of service by the adoption of these devices, either positively (to subscribe) or negatively (to unsubscribe)*"[272]. Moreover, the user of the connected speaker is locked into the user's initial choice by limited access to services that are "*often specific to a manufacturer. Content publishers should be able to access connected speaker environments on fair and equal terms*"[273].

Voice assistants are vertically integrated, but in most cases, they do not use an exclusivity strategy for applications when users implement their own configuration and integration process. Only Siri relies on a closed ecosystem that limits the choice of voice applications to Apple's related services. This business model has been challenged by the provisions of the DMA. The absence of exclusivity is in itself a strategy to avoid a new entrant in the voice assistant market using the exclusivity model to attract users, a practice that was employed in the field of video games

---

[270] European Commission, Final Report - Sector Inquiry, supra, point 459.
[271] ARCEP, Report supra, p.31.
[272] HADOPI / CSA, Study supra, p.74.
[273] HADOPI / CSA, Study supra, p.75.

between Sony and Microsoft[274]. Alexa lets users link the Amazon Firestick and the Amazon Echo speaker, giving access to all Amazon services. This is particularly true in the audiovisual field. But the user can choose YouTube and leave out the services offered by Amazon Prime. The same goes for the Google Home speaker, which acts as a remote control and allows users to control their TV by voice when paired with the Chromecast key plugged into the TV or built directly into the TV. Synchronization between the two devices is done by wifi, then via the Google Home application downloaded on the smartphone, through which users can add peripheral equipment and Google accounts[275]. However, there is nothing to prevent users from accessing the services offered by Amazon Prime. The strategy adopted in this market is one of a "colonization" of ecosystems which allows users, via the voice assistant or the connected speaker, to expand the initial ecosystem or even extend it to other ecosystems.

The main difficulty of discoverability lies in the device's role as an intermediary, which poses technical obstacles likely to reinforce the user's relationship with the voice assistant to the detriment of the services offered. Users must configure and integrate the technical specifications of the services they want to use on the voice assistant in order to use it, which assumes that third-party companies can guarantee the interoperability of their service with the technical system of the voice assistant. This makes it difficult for user companies to monitor the onboarding experience of their end-users as underlined by the European Commission[276]. In reality, users accessing third-party services will do so for the first time via the voice assistant interface, which does not allow third-party companies to access user data or metadata. Moreover, in most cases, the user has an account and a username specific to the voice assistant, which creates an additional barrier between the user and the user companies.

The European Commission further notes that pre-installation of applications constitutes a competitive disadvantage for content providers whose applications are not pre-installed, as uses have to take additional steps to access a service that is not included in the offering they have subscribed to. Conversely, pre-installed services enjoy higher levels of discoverability by users due to their visibility[277]. Discoverability is coupled with diversity when these pre-installed applications call up linear, random, or pre-constituted music streams in the form of a playlist. This is both a challenge and a risk. The proliferation of pre-installed applications locks users into the illusion of choice. However, pre-installing linear streams (and not just in the field of music) and promoting podcast aggregators are both worth thinking about, provided that the user has the option of easily uninstalling these services. Within this constrained framework, it is important to reconcile, with the help of algorithms, the personalization of results with the possibility of leading the user to other content. Personalizing choices is not an

---

[274] French Competition Authority, Decision No. 15-D-18 of 1 December 2015 regarding practices implemented in the video game sector.
[275] HADOPI-CSA, Study supra, p.75.
[276] Idem, point 480.
[277] Idem, point 449.

obstacle to diversity, and neither is configuring recommendation algorithms that can integrate all humanly conceivable variables (increasing and adding recommended content, adding random content, setting up hybrid recommendation systems)[278].

The more unique the answer, the more important it is to rank the content. Based on ARCEP's recommendations[279], the idea of promoting content diversity through discoverability implies that listing and ranking criteria used by voice application stores related to voice assistants must first be known by developers. Application stores are opaque about the confirmation and search optimization rules for these applications. As a result, developers have no visibility on the availability of their applications. Balanced information should be ensured in order to address the contractual imbalances highlighted by the Commission[280]. The Commission has found that the discoverability of voice applications is controlled by the major multipurpose voice assistant vendors, who do not disclose these rules to third parties. From the feedback provided to the Commission, one creative content service provider explains that, *"Integrated service providers have the short-term advantage of being easily retrievable via the integration on the voice assistant. However, in the long run, such integration makes these services dependent from the intermediation power of the voice assistant. Also, this intermediation happens with no algorithmic transparency, eventually forcing service providers to pay if they want their content to be visible to consumers"*[281].

User protection against the installation of malicious applications. The DMA challenges restricted access to voice applications and in particular the integration system (see above) and supports opening it up to third-party user companies. However, one point worth highlighting concerns issues related to user protection, especially when users access malicious applications. The technical integration system that brings interoperability into play can, among other things, protect the user in that applications could protect their security or the security of their voice assistant by exploiting software or hardware vulnerabilities, their personal data, or even protect them from counterfeiting. Therefore, either the application is malicious to the voice assistant or to the user. In most cases, applications are set up at the initiative of the user through an interface that can be misleading, hacking of the application editor, or through the use of malicious software. Depending on the choices set by the user, different services can be accessed by a voice assistant without the need for authentication. The Digital Regulation Expertise Network (PEReN - *Pôle d'expertise sur la régulation numérique*) provides insight into security issues related to the distribution of applications outside of operating system stores,[282] and it details existing protections against malicious applications. Their work does not specifically deal with voice applications, but the reasoning is applicable. An operating system should be

---

[278] On this subject, see the extensive work of V.-L. Benabou and J. Farchy, Rapport, supra; M. Foulonneau, V. Groues, Y. Naudet, M. Chevalier, "Recommandeurs et diversité : exploitation de la longue traîne et diversité des listes de recommandations", in G. Chartron, I. Saleh, G. Kembellec, Eds. *Les systèmes de recommandation*, Hermès, 2014, ch. 4.

[279] ARCEP, Report supra, p. 48.

[280] European Commission, Final Report - Sector Inquiry supra, points 482 and 483.

[281] European Commission, Final Report - Sector Inquiry, supra, point 481.

[282] PEReN, February 2022, No. 2, https://www.peren.gouv.fr/rapports/2022-02-18%20-%20Eclairage-sur-ouverture-os_FR.pdf

protected from vulnerabilities through a permissions system that tracks application behaviour and limits interface functionality to prevent attacks. From this point of view, the voice interface without the voice assistant could be reassuring, but only falsely so since it could also be turned into a black box. Application certification is therefore necessary, regardless of how applications are installed.

However, in its inquiry into connected objects used by the general public, the Commission stipulates that "*prior to being made available on a voice assistant, third-party voice applications must initially accept the voice assistant's terms and conditions, and get the approval of the voice assistant provider through its proprietary review process*"[283]. This integration/certification process is not considered relevant by smaller third-party user companies. In general, the process is cumbersome and largely one-sided. The certification process must involve third-party developers, even though integration would actively combat malicious applications. But the system is too centralized.

Certification process categories for integration with voice assistants

a. "Works with" programs or connected device certifications allow smart device manufacturers to make their smart home devices controllable through a voice assistant embodied on another device (i.e. smart speaker) or support (i.e. smartphone application). For smart home device manufacturers, these programs generally involve reviewing the voice assistant provider's documentation on API functionality for smart home device interoperability and developing a voice application following the voice assistant provider's requirements. The voice assistant provider then tests the developed integration to ensure a good user experience. If certification is granted, manufacturers can use "works with" logos or badges in their packaging or for online marketing. Examples of these certification processes are the "Works with Alexa", "Works with Google Assistant" and "Works with Apple Home kit" programs.

b. Built-in certification processes allow third party smart device manufacturers to support Amazon or Google's cloud-based voice assistants on their devices and gain the "Alexa built-in" or "Google Assistant built-in" badges. Apple, on the contrary, does not offer built-in certifications, as Siri can only be built-in on Apple's own devices. Through built-in integration, users can access all the services supported by the thirdparty voice assistant from the manufacturer's smart home device, including smart device control and access to third-party services through voice applications. In this regard, built-in solutions are different from works-with programs, which are limited to smart device control functionalities. Built-in certification processes involve a thorough testing of the smart devices in which the voice assistant will be embodied by the voice assistant provider. On a technical level, manufacturers need to integrate a built-in microphone and comply with the required hardware requirements specified by the voice assistant providers.

c. Development of voice applications, such as "Alexa skills", "Google actions", "Siri' shortcuts" or "Capsules" for Bixby run on the voice assistant for which they have been

---

[283] European Commission, Final Report - Sector Inquiry supra, point 183.

developed. Voice applications make consumer IoT services accessible to the user via voice assistants. From a technical perspective, voice assistant providers make available SDKs to build custom voice applications. Developers must follow the voice assistant guidelines on how to write an application and also comply with the natural language understanding ("NLU")35 rules of the voice assistant. The voice application is then submitted for certification before it is published and made available through the voice assistant.

Source : European Commission, Commission staff working document. Preliminary report – sector inquiry into consumer Internet of things, 2021, SWD(2021) 144 final, point 189.

PEReN also offers an application auditing system to analyze application behaviour before installation[284]. The aim is to establish a trusted third-party mechanism for the user. Application stores audit the applications they distribute, but this audit can be revoked by the auditor or the application developer. This audit work could be done directly by third-party application stores or by auditing organisations independent of the distribution, provided that they provide follow-up monitoring. The security of voice assistants legitimizes both the pre-installation of some applications (provided that the user can uninstall them) and also the certification mechanisms, provided that the technical environment and that of the actors are more open. At the same time, the user's role is to ensure that voice assistants are used on up-to-date and secure equipment so as to limit any maliciousness, which could constitute a significant breach in the protection of personal data, such as banking identity.

The voice and only the voice. The use of voice is a structural constraint; the user must keep in mind that the interaction is carried out only by voice. Any tools contributing to informing users about the voice assistant and the data processing and ecosystem learning processes alter the voice interaction by making it less seamless. However, these mechanisms, which are used to protect personal data, acquire consent, and protect the user's freedom of choice, are decisive. There is still considerable innovation underway in the rapidly expanding "voice market", with the imminent arrival of chatbots whose powerful algorithmic systems suggest that they will be able to perceive the mental and emotional load of the user through the collection of data in the context of voice interactions. That is why it is essential not to limit the issue of diversity to the discoverability of content, but to imagine that greater transparency could be achieved in particular through a "data-based" approach to regulation advocated by ARCEP, so that the level of information would be comparable for users, individuals, and professionals[285]. If information is provided at all levels, the user can move from one ecosystem to another within the framework of data portability and technical interoperability, and it would also be possible to lift restrictions linked to

---

[284] PEReN, *Shedding light on… n°2// Mobile applications: what are the security issues for their distribution outside OS stores?* February 2022, p. 4. https://www.peren.gouv.fr/en/actualites/2022-02-18_eclairage_sur_os/
[285] ARCEP, Report supra, p. 46 and following.

the integration of voice assistants or prioritize content and services made available in a random and alternative way (see below).

Diversity in supply and demand. This is an interesting approach and has been the subject of much discussion in the context of Europe, where, from the user's level, the ecosystem approach would help create interdependence between markets, services, algorithms, data, and actors. Looking at diversity from the perspective of supply limits the choice of tools for designing it. Thus, considering diversity within the framework of demand helps give a comprehensive view of all the actors and gives users power to act.

C - The user: a new actor for pluralism and diversity

For there to be communication between two people, there must be a transmitter and at least one receiver. This two-way communication presupposes that these two actors, in order to be able to interact, have the same prerogatives. However, the receiver has long been considered passive and largely ignored by legislators.

Towards a user's right to act. The wording of Article 11 of the Declaration of the Rights of Man and of the Citizen (DDHC), a legacy of the French Revolution, is remarkably precise. Its scope is the sender and the sender's freedom, rights, and obligations. It does not target the receiver, who is considered to be passive, even if the universal character of the DDHC extended the freedom of communication to all means proceeding from the externalization of thought. The Constitutional Council considers that Article 11 is applicable to all electronic communications[286]. The receiver has never been defended as a legally protected interest. Communication legislation was built with this imbalance, based on the protection of the receiver by the sender, with the communications professional in the dominant position. The European Convention on Human Rights, in its Article 10, while recognizing the freedom of reception and access to information, does not provide greater justiciability for the receiver. The receiver's status is neutralized by the reference made to the public, or to the public's right to information. The public interest remains among the interests to be preserved, by the issuer, but also by the judge[287]. In June 2020, the Council once again interpreted this Article 11 by recognizing, alongside the freedom to access digital services, the freedom to express oneself, especially on social media[288]. The decision censured a device because of the infringement it caused to this freedom of expression through the risk of over-censorship. It thus considered, for the first time, the active role

---

[286] E.g., Cons. const. 3 March 2009, No. 2009-577 DC, recital No. 3, D. 2009. 884, viewpoint A. Levade, and 2010. 1508, obs. L. Gay; AJDA 2009. 617, tribune P. Wachsmann; RFDA 2009. 580, chron. A. Roblot-Troizier and T. Rambaud; *Constitutions* 2010. 109, obs. A. Barilari, and 238, obs. M. Disant.
[287] On these points, K. Favro, *Droit de la régulation des communications numériques*, LGDJ, coll. Systèmes, 2018.
[288] Cons. recital, 18 June 2020, No. 2020-801 DC.

of users on the basis of Article 11 of the DDHC. These users, for whom the General Data Protection Regulation (GDPR) and the French Data Protection Act (LIL) used to recognize a certain empowerment within the framework of the protection of their personal data, are now recognized with the same prerogatives as regards freedom of expression. Guaranteeing end users the freedom to express themselves on social media has the consequence of subjectivizing their role. Consequently, the freedom of choice that was guaranteed by the Constitutional Council in the 20th century now takes on its full meaning. This makes it possible to go beyond initiatives aimed at promoting the diversity issue only from the perspective of platforms by considering the diversity of demand. The user therefore has a role to play in characterizing diversity. The aim is to guarantee end users a right to action so that they can make their own choices.

Several levels of action are necessary considering the network effect and the effects of scale on the massive use of data so that the user can be empowered. First, the user's capacity to act on data and become a freely choosing agent needs to be considered. This would presuppose intervention from regulators (1). Beyond that, other means should be considered to rethink user interactions with the voice assistant provider (2).

1. Empowering the user through data

Regulation by data. ARCEP, in the context of its study on devices, has formulated a number of recommendations so that devices can contribute to making the Internet more open by improving the ability of users to act, particularly in the choice of devices and their ability to change them[289]. The authority proposes regulating through data as a means of guaranteeing information so the regulator can "*make objective diagnoses*". This would apply to information for the user, which would be a pillar of empowerment in an ecosystem that is biased due to the ability of structuring actors to capture the voice assistant market without effectively guaranteeing the end consumer access to all the services offered : "*More and better information could help rectify the behaviour of equipment suppliers and operating system providers, by giving users the power to make informed choices*"[290]. ARCEP proposes to clarify the scope of the open Internet by establishing a principle of freedom of choice of content and applications regardless of the device. As the nerve center of the ecosystem, data contains all the information useful to market actors, including users, thereby contributing to the transparency, fluidity, and contestability of the market. Indeed, to open up ecosystems and their core platform services, two actions must be taken in concert, conditioning the implementation of concrete ex ante regulation tools dedicated to empowering the user.

The first action is taken by the regulator and is based on its ability to collect information from device manufacturers and operating system publishers. Available

---

[289] ARCEP, Report supra, p.47.
[290] Ibid.

information and data are analyzed with respect to the extent that they are lacking for all actors, and not just for regulators. Voice application vendors have little visibility into the effects of application availability on operating systems, which includes usage data and system performance data. While ARCEP advocates the publication of these data, it also encourages academic and independent research in this field. The ARCEP can reinforce the regulator's actions. Article 40 of the DSA provides for oversight of very large online platforms and search engines by vetted researchers affiliated with a research organisation. However, the DMA has no such provision. Even if both texts have a considerable impact on data management through their multi-disciplinary approach, the identification, understanding, and evaluation of systemic risks (justifying researchers' access to the data of services and products proposed by the gatekeepers) remains the fundamental issue of the DSA (Articles 34 and 35 of the text). However, such an approach would have been legitimate to support the contestability of markets. A reflection on the expertise of the regulator must be carried out in order to consider a more detailed regulation of the different ecosystems and more particularly with respect to access to the data collected and processed by voice assistants.

PEReN is a department whose competence is national in scope (France) should develop systems of action to gain visibility in this area. Through the missions assigned to PEReN, it embodies a center of scientific expertise in data seized by the services of the French state and the regulatory authorities which make use of it (ARCEP, ARCOM, CNIL, French Competition Authority, French Ministry of Culture, French Ministry of Economy, finance, and Industrial and Digital Sovereignity). It intervenes both as a support service within the framework of agreements that integrate the regulatory specificities of the relevant administrative authorities on the capacity to collect data. It also intervenes as an expert within the framework of research work "*commissioned by French state services, by carrying out studies of an exploratory or scientific nature: PEReN then acts as a research centre*" [291].

In particular, in 2021, PEReN analyzed, in the framework of a partnership with HADOPI, the supply and consumption of videos offered on YouTube channels[292], which makes it possible to have sufficiently detailed elements of analysis to characterize the diversity of content made available by this actor, which, moreover, is accessible via voice assistants. Similarly, PEReN provided insight into AMP (Accelerated Mobile Pages), introduced by Google in 2015, designating a set of tools that work to accelerate the display of web pages on mobile devices in a way that avoids ad blockers, which may have been useful to news publishers[293] or, in 2021, insight into interoperability[294]. In this context, PEReN is able to analyze a market and perceive its evolutions. Its REGALIA division is also specialized in the study of algorithms. Here,

---

[291] https://www.peren.gouv.fr
[292] https://hadopi.fr/sites/default/files/2021-12/2021_12_15_rapport_YouTube_PEReN.pdf
[293] PEReN, PEReN, *Shedding light on…N°5 - Google's MPAs: from doping to weaning?*, September 2022, https://www.peren.gouv.fr//rapports/2022-10-17%20-%20Eclairage-sur-AMP_FR.pdf
[294] https://www.peren.gouv.fr/rapports/2021-10-08%20-%20Eclairage-sur-interoperabilite.pdf

it has provided expertise on the virality of content at the request of ARCOM, on the minimization of data obtained in portability on behalf of the CNIL, on the audit of algorithms in low transparency, or on an experimental device to automate the analysis for services that present their information only through their applications on smartphones. Some of these expert reports are confidential, and all of them were initiated by the administrative authorities, and they do not have the visibility or spontaneity of academic research. However, the regulator's expertise must be based on detailed and cross-sectional data in order to have the right level of information and be able to intervene ex ante and asymmetrically on the market. The conclusions of ARCEP's study on devices call for this data collection "*to stimulate the availability of a wider selection of available content and services*"[295], and consider collecting user reports, following the example of what has been undertaken in the field of electronic communications, in order to map the risks generated by the practices of market actors. This would allow the regulator to filter the user's expectations, although it would not resolve them but essentially document them. Similarly, by aggregating information from third-party actors, the regulator can develop qualitative or rating comparison tools for the leading voice assistants to improve oversight of the practices of manufacturers and operating system vendors[296]. The objective is to allow for the emergence of ten thousand third-party actors (SMEs and start-ups) dependent on the structuring actors. Based on this position, the regulator must be given the possibility of intervening in the framework of dispute resolution mechanisms to deal with disputes between structuring actors and third-party user companies. From another perspective, Regulation 2019/115 provides for the settlement of disputes between platforms and their professional users, whether it is the establishment of extrajudicial mechanisms (internal complaint handling system and mediation) or the possibility enshrined for the benefit of various actors to initiate legal proceedings (representative association bodies, public bodies)[297]. The procedure recommended by ARCEP is based on a regulatory approach, not on an approach through restrictive competition practices.

ARCEP, Recommendations

"An expert regulator could be given the power to settle disputes related to internet openness and devices, including those regarding their operating systems, browsers, and app stores. In particular, there can be an imbalance in the relationship between certain content and service providers and certain device manufacturers. It could thus be advisable for content and service providers to have a path for swift and efficient recourse when a dispute arises over the conditions governing their access to devices, whether through operating systems, browsers, or app stores. A dispute settlement procedure could be introduced to arbitrate, on a case-by-case basis, disagreements between professional actors regarding:
- the indexation and ranking criteria used by app stores,
- operating system showcasing certain content and services,

---

[295] ARCEP, Report supra, p.55.
[296] ARCEP, Report supra, p.47-48.
[297] Regulation 2019/1150, Articles 11-14.

- third-party search engines' access to app stores,
- third-party application stores' access to operating systems,
- application access to operating systems' APIs".

Source: ARCEP, *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open internet: Report on their limitations and proposals for corrective measures, Feb. 2018 p.57.*

The second action, which correlates with the first one, supposes that the information acquired by the regulator is put at the service of the user so that the user understands the functioning of the voice assistant "from start to finish" in order to be able to make informed choices. This mainly concerns the practices of developers or the criteria that make up the offering in application stores. In reality, this deals with schematically informing users of all the practices likely to call into question an open Internet according to the recommendations of the ARCEP, but also of proposing tools that will give users the power to act. This action aims at transferring to users the guarantee of pluralism with respect to content, in addition to the existing mechanisms which until now have been based exclusively on content publishers.

2. Freedom of choice and interaction with the digital service

With regard to user interactions with the digital service, the exercise of the user's freedom of choice must be carried out in two stages. First of all, the user must be provided with all the information necessary for an exercise of choice, which means that the choice cannot be the result of manipulation or dark patterns (a). Second, the user of the voice assistant must be able to exercise choice by acting on the technical infrastructure, in particular by configuring the tool according to user preferences (b). Finally, users must be able to migrate to new services (c).

a. Freedom of choice on the basis of user information

Ensuring the user's understanding of the operation of the voice assistant and associated services in order to guarantee diversity through demand implies first of all that the user's choice should not be misled through deceptive and manipulative choice interfaces; it also requires that the user be informed of content ranking methods.

Deceptive choice interfaces (dark patterns). As the CNIL has noted, "*the design of human-machine interfaces, far from being a simple question of ergonomics, involves crucial issues in terms of capacity to act and the configuration of choice*

possibilities"[298]. However, design techniques can be based on the exploitation of cognitive biases operating automatically when an individual interacts with an information system and "*elements or mechanisms can be put in place in interfaces to influence users' decisions in a direction they would not necessarily choose if they were faced with a fair and transparent interface*"[299]. Guaranteeing the user's freedom of choice with respect to a voice assistant or chatbot therefore implies paying particular attention to the choice interfaces. This point is essential given the specificity of voice assistants: on the one hand, given their particular interface (voice interactions, single response and absence of screen, link between the voice assistant and its ecosystem as described above); on the other hand, because of the possibilities offered to the voice assistant provider to personalize and optimize interactions with its users through the algorithmic analysis of the impact of the design on the user of the service based on the analysis of a large mass of usage data (metadata and personal data) In this respect, the CNIL notes that "*these new interaction methods are intended to be revolutionary because of the spontaneous understanding that the user would have and the progressive learning curve that they offer*" and even "*open the door to experiences that are more and more based on emotions*"[300].

An important reflection is now underway in order to sanction or prohibit the use of misleading or manipulative choice interfaces as this can have detrimental effects on consumers and competition. In particular, reference should be made to four important papers published in 2022 by the Federal Trade Commission[301], the UK Competition and Markets Authority[302], the European Commission[303], and the OECD[304], which echo previous studies conducted in France by the CNIL[305], in Norway by the Forbrukerrådet (Consumer Council)[306], and more recently in the Netherlands by the Netherlands Authority for Consumers & Markets[307]. These papers present various types of issues or even illicit practices that demonstrate, for instance, choice structures (e.g., altering the presentation of choices for self-preference), choice information (e.g., making information difficult to understand or giving too much information), and choice pressure (e.g., using user habits, time pressure, or trust messages). The OECD notes that "*the most common dark patterns on websites and apps are framing (pre-selecting or visually prioritizing choices, hiding information, or hiding ads), creating a sense of urgency (through potentially misleading scarcity cues), generating social proof (through potentially misleading popularity cues), forcing registration or disclosure of information, inducing choice, or making it difficult to cancel or opt-out*"[308].

---

[298] CNIL, *La forme des choix. Données personnelles, design et frictions désirable*s, Cahier IP No. 6, 2019, p. 7.
[299] E. Harry, "Dark patterns : quelle grille de lecture pour les réguler ?", *LINC*, 2 September 2019.
[300] CNIL, report supra, p. 9.
[301] Federal Trade Commission, *Bringing dark patterns to light*. Staff Report. AN FTC Worhshop, September 2022.
[302] Competition & Markets Authority, *Online Choice architecture. How digital design can harm competition and consumers*. CMA 155, April 2022.
[303] F. Lupianez-Villanueva & al., *Behavorial study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation*. Final report, European Commission, DG Just, April 2022.
[304] OECD, *Dark commercial patterns*, OECD Digital economy papers, October 2022, No. 336.
[305] CNIL, *La forme des choix. Données personnelles, design et frictions désirable*s, supra.
[306] Forbrukerrådet, *Deceived by design - How tech companies use dark patterns to discourage us from exercising our rights to privacy*, 2018.
[307] ACM, *Protection of the online consumer. Boundaries of online persuasion*, guidelines, 2021.
[308] OECD, *Dark commercial patterns*, OECD Digital economy papers, October 2022, p.5.

Taxonomy of the main commercial dark patterns

Forced action
A forced action seeks to force the consumer to do something in order to access a specific feature such as registering or being forced to disclose more personal information than desired.

Interface interference
Design manipulation aims to elicit specific consumer actions conducive to online commerce through how information is organised, e.g., hiding information, pre-selecting options, creating a false hierarchy, using intentional ambiguity.

Nagging
Nagging involves repeated requests to the consumer to do something favorable to the company, such as turning on notifications or location tracking, and can thus exploit the consumer's limited willingness or time.

Obstruction
Obstruction-related dark patterns aim to make a task or interaction more difficult than it needs to be, with the intention of deterring an action, and can therefore exploit the consumer's inertia, limited willingness, or time.

Sneaking
Sneaky dark patterns seek to hide, disguise, or delay the disclosure of information relevant to the consumer's decision, particularly with respect to cost, and may exploit consumers' limited attention, default bias, anchoring effect, or sunk cost fallacy.

Social proof
These interfaces attempt to trigger a decision based on observations of other consumers' behaviour, and can therefore exploit the bias of social pressure or approval.

Urgency
Urgency imposes a temporal or quantitative limit, real or false, to an offer in order to push the consumer to make a purchase, thus exploiting the scarcity heuristic.

Source: *OECD, Dark commercial patterns, Digital economy papers, October 2022, p. 10*

Sanctions for deceptive choice interfaces. These practices can be sanctioned on the basis of unfair commercial practices within the meaning of Directive 2005/29, known as the PCD[309], and, in France, of Articles L. 121-1 and following of the Consumer Code. In this regard, it should be noted that the updated December 2021

---

[309]Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive').

SCI guidance specifically targets dark patters to refer to "*malicious nudging, generally incorporated into digital design interfaces. Dark patterns could be data-driven and personalised, or implemented on a more general basis, tapping into heuristics and behavioural biases, such as default effects or scarcity biases*"[310]. For example, an unsubscribe procedure is considered an unfair practice if it is too complex, or if emotional messages are sent to the subscriber ("*We are sad to see you go*", "*Here are the benefits you will lose*"), or if visual interference is used to encourage subscribers to keep their subscription rather than to terminate it by means of various images. This explains the recent evolution of the procedure regarding the *Prime service* offered by Amazon. The European Consumer Organisation (BEUC) filed a complaint against the company with the European Commission, citing numerous obstacles to unsubscribing from the service (complex navigation menus, convoluted wording, confusing choices, and repeated soft prompts). Since then, Amazon has committed to changing its interface to simplify its unsubscribe process; it should now only take two clicks with a prominent and unambiguous "cancel" button, so that unsubscribing from a service is as easy as subscribing to it.

Misleading choice interfaces may also be subject to sanctions for violation of personal data law, as the CNIL has acknowledged in various recent decisions. Such practices are contrary to the principles of transparency, guaranteeing informed consent, and the principle of privacy by default by guiding the choices of data subjects without their knowledge, as the CNIL has recognized in various decisions[311] or, more recently, the EDPB in its guidelines on dark patterns in social networks[312].

EU law and the principle prohibiting deceptive choice interfaces. EU law reforms are now considering going further by enacting a principle prohibiting deceptive and manipulative choice interfaces. For example, Article 13 of the DMA explicitly prohibits circumventing the obligations imposed on the operators covered by the text - the gatekeepers - through conduct, "*contractual, commercial, technical nature, or of any other nature, or consists in the use of behavioural techniques or interface design*". In addition to this "anti-circumvention rule", there is a prohibition against deteriorating the conditions or quality of service and making it excessively difficult for users to exercise their rights or make choices, including by offering choices in a biased manner, or by using the structure, design, function, or mode of operation of a user interface or related part to disrupt their autonomy, decision-making, or free choice. It should be noted, however, that this text does not impose a general prohibition on deceptive interfaces but only concerns the scope of application of the DMA. On the other hand, more IoT operators will be affected by this prohibition when the proposed Data Act regulation is adopted. Article 6.2.a of this text provides that third parties receiving data at the

---

[310] European Commission, *Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market*, 29 Dec. 2021, 2021/C 526/1, point 4.2.7.
[311] CNIL, Deliberation 21 January 2019, No. 2019-001; Deliberation 31 December 2021, No. 2021-023 and No. 2021-024; Deliberation No. 2022-020, 10 November 2022, Discord INC - Adde, CNIL, *La forme des choix* supra., p. 39 and following.
[312] EDPB 3/2022, *Dark Patterns in social media platform interfaces: How to recognize and avoid it*, 2 May 2022.

request of the user shall not "*coerce, deceive or manipulate the user in any way, by subverting or impairing the autonomy, decision-making or choices of the user, including by means of a digital interface with the user*" [313].

The proposal for a regulation known as the "AI Act" appears to be much less ambitious, given the current state of negotiations. The AI Act provides that, among the prohibited practices (which are subject to various layers of obligations depending on the risk generated), the only uses that are covered are those that "*deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person physical or psychological harm*" (Article 5.1.a). However, the very restrictive conditions set by this text will considerably limit its scope, unless changes are made prior to its adoption[314]. This is nevertheless a major issue, especially in the context of chatbots and the potential exploitation of emotional data for the purpose of manipulating their users[315]. In this regard, we refer to the analyses of the Council of Europe, which, in its Declaration on the manipulative capacities of algorithmic processes, noted that, "*Contemporary machine learning tools have the growing capacity not only to predict choices but also to influence emotions and thoughts and alter an anticipated course of action, sometimes subliminally*"; "*Fine grained, sub-conscious and personalised levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions*"[316].

Ranking and user information. As far as diversity is concerned, one of the major challenges in terms of user information concerns the ranking of voice applications available to the user and therefore the cultural content that can be accessed via the voice assistant. This situation is all the more problematic as the user rarely has access to several app stores, but is limited to the one offered by the voice assistant. Users should therefore be informed of how content is ranked so that they can adapt their request and enrich it with the necessary details to orientate the response according to their interests.

Ranking, recommendations, preference. As noted, ranking can be seen as a way to promote certain content. To that end, the CSPLA mission report on recommendations focused on content ranking as a prerequisite in recommendation mechanisms. Article 3.s. of the DSA defining recommendation systems sheds a useful

---

[313] See also Article 25 of the DSA: "Providers of online platforms shall not design, organise or operate their online interfaces in a way that deceives or manipulates the recipients of their service or in a way that otherwise materially distorts or impairs the ability of the recipients of their service to make free and informed decisions".

[314] On this point, see S. Vergnolle, "L'illusoire interdiction des pratiques manipulatoires dans le projet de règlement sur l'intelligence artificielle", *Un droit de l'intelligence artificielle : entre règles sectorielles et régime général*, eds. C. Castets-Renard and J. Eynard, Bruylant, 2023, p. 437.

[315] See in particular CNPEN, *Ethical issues of conversational agents*, opinion No. 3, Nov. 2021, p. 7 - R. Chatelier, "Captation des émotions : comment vous le direz pourra être retenu contre vous… ", *LINC* 4 April 2018.

[316] Council of Europe (2019), *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes* points 8&9- Adde, N. de Marcellins-Warin, F. Marty, E. Thelisson, and T. Warin, "Artificial Intelligence and Market Manipulations: Ex-ante Evaluation in the Regulator's Arsenal", *RIDE* 2020/2, t. XXXIV, p. 203, spec. p. 215 and following.

light in its Recital 70, which precises that, "*A core part of the online platform's business is the manner in which information is prioritised and presented on its online interface to facilitate and optimise access to information for the recipients of the service. This is done, for example, by algorithmically suggesting, ranking and prioritising information, distinguishing through text or other visual representations, or otherwise curating information provided by recipients*". Ranking is a step in the recommendation mechanism, which does not disappear from the 'creation' of a response". This recommendation mechanism is not essential because the prioritisation is not known to the user. It refers to a presentation architecture that will highlight given content without necessarily recommending content in the true sense of the term. "*In other words, a platform deciding to favour a certain way of presenting content does not necessarily qualify as a 'recommendation'*"[317]. In our analysis, the content put forward by voice assistants can sometimes be based on self-preference mechanisms that cannot be considered to be recommendations. On the other hand, in all these hypotheses, upstream content ranking remains the common denominator, which can raise various questions when rankings lead to content by the structuring actors taking precedence.

Informing users of algorithmic ranking. Many European texts address the issue of product ranking information, but they do not necessarily target users. The question of how products and services are ranked is crucial in platform ecosystems with respect to the platforms themselves, user companies wishing to promote their products, and users.

To that end, the so-called P2B Regulation[318] provides a definition of the term ranking as "*the relative prominence given to the goods or services offered through online intermediation services, or the relevance given to search results by online search engines, as presented, organised or communicated by the providers of online intermediation services or by providers of online search engines, respectively, irrespective of the technological means used for such presentation, organisation or communication*". However, this definition is aimed at relationships between professionals in order to protect user companies by preserving business secrecy. Consideration of commercial interests should not preclude disclosing the key parameters that determine ranking. The aim is not to disclose detailed insight into the ranking mechanisms and algorithms of relevant providers[319]. The regulation requires intermediation service providers to describe in their terms and conditions the main parameters that determine the ranking, as well as the reasons justifying their importance in relation to the other parameters, with the objective of informing user companies. In the case of online search engines, information about the main parameters must be stated in a clear and understandable way. The transparency

---

[317] V.-L. Benabou and J. Farchy, *Les dispositifs de recommandation des œuvres audiovisuelles et musicales sur les services en ligne*, final report presented to CSPLA, 2021, p.16.
[318] Article 2.8, Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services.
[319] See Recital 23 of Directive (EU) 2019/2161 and Recital 27 of Regulation 2019/1150.

requirement, even if it benefits the consumer, is only motivated by the interest of user companies; it provides "*a general description of the main ranking parameters and allows companies to appreciate their degree of relevance and have a good understanding of the functioning of ranking in the context of the use of intermediation services or search engines. However, it is not necessary for them to have precise and detailed knowledge of how the ranking mechanisms function or of the algorithms*"[320].

The regulation may benefit the consumer/user, but users and consumers are not its target. Users may legitimately expect search request results during human/machine interactions to be natural, or at least relatively neutral. However, it is possible for content providers to include paid advertisements in the results or to prioritise some products because of the direct or indirect payments received from the professionals selling these products. Therefore, Article 3.1.b of the Directive 2019/2161[321] provides a definition of ranking that takes the interaction with the consumer into account: "*the relative prominence given to products, as presented, organised or communicated by the trader, irrespective of the technological means used for such presentation, organisation or communication".* This article should be read in conjunction with Recital 19 of the Directive which precises this includes "*resulting from the use of algorithmic sequencing, rating or review mechanisms, visual highlights, or other saliency tools, or combinations thereof*". One of the contributions of this directive results from Article 3.4.b, which establishes a specific information requirement regarding the main ranking parameters: "*When providing consumers with the possibility to search for products offered by different traders or by consumers on the basis of a query in the form of a keyword, phrase or other input, irrespective of where transactions are ultimately concluded, general information, made available in a specific section of the online interface that is directly and easily accessible from the page where the query results are presented, on the main parameters determining the ranking of products presented to the consumer as a result of the search query and the relative importance of those parameters, as opposed to other parameters, shall be regarded as material. This paragraph does not apply to providers of online search engines as defined in point (6) of Article 2 of Regulation (EU) 2019/1150 of the European Parliament and of the Council*".

The information obligation in question is still too limited. It already assumes a consumer request. This information requirement applies only to professionals who allow consumers to search for products offered by other third-party professionals or consumers, i.e., online marketplaces and comparison tools. The obligation to inform also applies in the event that the professional allows searches to be made on an online interface by voice command, instead of typing the query. In this context, the information must be made available for review on the professional's

---

[320] G. Loiseau, "Le règlement Platform to Business", *CCE*, No. 7-8, 2020, study 13.
[321] Directive 2019/2161 of 27 November 2019 of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council.

website/application "*in a dedicated section of the online interface*". Even if it interferes with the voice interaction, the information should be available on a dedicated website or online application that is easily accessible from the page on which the offer or voice command is made available. The information must then be provided in a clear and comprehensible manner and in a form suitable for the remote communication technology. It is also specified that this information must be located in a dedicated section of the online interface that is directly and easily accessible from the page on which offers are presented. On the other hand, it does not apply to consumers who make choices and search for their own product offerings outside of those suggested by the structuring platforms without entering a query, nor does it apply to online search engines, as defined by Regulation 2019/115. In reality, if there is no request from the user/consumer, and if there is a default interface, the directive will not apply.

The level of information in interactions with users. Because the target here is users, information must be clear and concise. For the same reason, the Directive's rules do not require an explanation of the "reasons" for the relative importance of key ranking parameters, as the P2B Regulations do. While the information must remain general, as in the regulation, by providing the main parameters determining the product ranking, it must inform the consumer of the relative importance of these parameters in relation to the other parameters. However, Recital 23 of the Directive specifies that the description of the default ranking parameters can remain general and should not be dependent on individual requests. The two texts in question require platforms to inform their user companies by means of information included in the platform's general terms and conditions during transactions between companies, or to ensure that this information is available during the pre-contractual phase. The aim is to satisfy the transparency requirement before the parties go any further in the process.

Ranking criteria. These new rules ranking transparency are similar in both texts presented. Recital 22 of the Directive perfectly summarizes the intention, stating that, "*Parameters determining the ranking mean any general criteria, processes, specific signals incorporated into algorithms or other adjustment or demotion mechanisms used in connection with the ranking*". Under the P2B Regulation, intermediation services may provide more detailed information to their user companies in order to explain what the ranking parameters are for the consumer. In order to clarify ranking criteria and further compliance with the transparency requirement, the Commission has published guidelines specifying the [322]application of the P2B regulation. These guidelines provide useful food for thought[323].

DMA Contribution. The P2B Regulation and the UCP Directive merely impose a reinforced transparency obligation on platforms, while the DMA prescribes more

---

[322]European Commission, *Guidelines on transparency in ranking, in accordance with Regulation (EU) 2019/1150 of the European Parliament and of the Council* 2020/C 424/01, 8 December 2020.
[323]Communication from the Commission Guidelines on transparency in ranking, in accordance with Regulation (EU) 2019/1150 of the European Parliament and of the Council 2020/C 424/01

restrictive obligations, even if they only concern gatekeepers. The approach of these different texts does not follow the same logic. The focus of the DMA is on this dual role of intermediating core platform services with third-party user companies, but also directly safeguarding their products and services[324]. These actors question the contestability of their products and services, thus reducing user choice. Article 2 of the DMA considers the issue of ranking with respect to restrictive competition practices and specifically with respect to self-preference, specifically targeting voice assistants. This article then defines ranking as "*the relative prominence given to goods or services offered through online intermediation services, online social networking services, video-sharing platform services or virtual assistants, or the relevance given to search results by online search engines, as presented, organised or communicated by the undertakings providing online intermediation services, online social networking services, video-sharing platform services, virtual assistants or online search engines, irrespective of the technological means used for such presentation, organisation or communication and irrespective of whether only one result is presented or communicated*". The text states that core platform services must not give themselves differential or preferential treatment in rankings, "*be it by legal, commercial, or technical means*". The conditions for ranking must be fair, covering all forms of relative priority, even those resulting from technical constraints, and in particular voice responses are targeted[325]. Gatekeepers shall not treat their own services or products, or those of third parties belonging to the same company, more favourably through rankings or other parameters than similar services or products from a third parts; and they must apply transparent, fair, and non-discriminatory ranking conditions for third-parts products and services. Therefore, core platform services must apply transparent, fair, and non-discriminatory ranking conditions to products and services similar to their own and provided by user companies. Although service discoverability may depend on rankings, simple findability is not assured[326]. In general, the contractual agreements that govern the visibility of some content have come under criticism because the negotiations are multinational and local services are generally excluded[327]. In the context of its sectoral inquiry, the European Commission recommends parity parameters allowing each user company to associate its own content with that of the leading actor, which is more visible, while also retaining the possibility of improving the visibility of content with, for example, the addition of a logo, which would ensure that the promotion of content is comparable, or at least equal. It should be noted that in this industry inquiry and with an eye towards strengthening this parity, one respondent "*suggested that leading providers of consumer IoT technology platforms may be able to exploit applicable algorithms to their own advantage, for example by associating specific search terms to their first-party content so that it appears higher up in the ranking*"[328].

---

[324] Recital 43 of the DMA.
[325] V.-L. Benabou and J. Farchy, Report supra, p. 57.
[326] European Commission, Final report - Sector Inquiry supra, point 466.
[327] Idem, point 467(a).
[328] Idem, point 467(b).

The objective is to allow users, through appropriate information, to dissociate the SEO and ranking of voice applications by giving them the opportunity to interact with all app stores and make choices. The aim is to take into account how users access their services. If they are accessing services through voice assistants, it would be appropriate to "*take into consideration the solution used by their service users to access other information about the service*"[329]. However, when it comes to audiovisual content, when the voice assistant features other embedded services, the difficulty may lie in the existence of remote control buttons that allow direct access to certain services or to a voice assistant. In this case, rankings are neutralized by purely material considerations ensuring a visual preference to such or such service due to the interface[330]. Following the recommendations of ARCEP in its study on device neutrality, it is up to the informed user to "*circumvent app stores' ranking algorithms and play a more active role in configuring the selection criteria applied to the applications available on app stores*"[331]. The user would then be able to override the content ranking governing responses.

b. Freedom of choice conditioned by potential user action on the system

Beyond this information, users must also be able to act on the interface with which they interact. This implies a particular interest in the different facets of the default settings, which may concern the choice of the voice assistant per se, the installed applications, or the applications called up by the voice assistant.

Default settings. Default settings, whether pre-installed or set as defaults, can serve the interest of users of digital services by helping them avoid a large number of active choices that would require them to expend considerable time and resources. On the other hand, default settings can be seen to be misleading if they seek to distort behaviour. In this sense, the Commission's guidelines on the UCP Directive emphasize that these practices "*have a significant impact on the transactional decision of an average consumer. Traders could not only influence consumers to take certain actions, but also take specific actions in their place, for example by using pre-ticked boxes, including to charge for additional services, which is prohibited under Article 22 of the [Consumer Rights Directive]*"[332]. Default settings may also distort competition, particularly when they allow an operator to exploit market power to retain or acquire new customers and favour its related services over third-party services[333].

---

[329] European Commission, Guidelines supra, point 119.
[330] European Commission, Final report - Sector Inquiry supra, point 469.
[331] ARCEP, *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open internet: Report on their limitations and proposals for corrective measures*, Feb. 2018, p. 55.
[332] European Commission, *Guidance on the interpretation and application of Directive 2005/29/EC of the European Parliament and of the Council concerning unfair business-to-consumer commercial practices in the internal market*, 29 Dec. 2021, 2021/C 526/1, point 4.2.7.
[333] V. Competition & Markets Authority, *Online Choice architecture, How digital design can harm competition and consumers*. CMA 155, April 2022, point 1.21 - comp. the Android case: Trib. EU, 14 Sept. 2022, case T-604/18, Google and Alphabet v. European Commission.

In order to ensure contestability and fairness in digital markets, the DMA now regulates these practices by imposing ex ante obligations on gatekeepers through[334] various provisions that can be classified into three categories, several of which extend the recommendations already made by ARCEP in its report on device neutrality[335]. Some provisions, like Article 6.3 of the DMA (which requires that the user be allowed and enabled to easily change default settings on their virtual assistant), should allow for end users to easily change default settings. Others tend to require the use of default initial choice screens to force end users to make active choices, which again will be required for voice assistant services under Article 6.3 of the DMA. Finally, the text intends to promote end-user choice in a general way, in order to encourage the switching of service providers and multi-hosting. To this end, Article 6.4 should allow for the installation of third-party application stores or applications that may also be offered as defaults, provided that they do not compromise the integrity of the device or system; Article 6.6 requires the user to be able to switch from one service to another without technical or other restrictions, and to subscribe to different services, while Article 6.3 allows the user to uninstall applications from the system if they are not essential to its operation.

Effectiveness Principle and setting configuration. These measures must be read in the light of the principle of effectiveness imposed by Article 8 of the DMA, since this text requires the gatekeeper to adopt measures to ensure compliance with the obligations imposed under Articles 5 and 6 and, more generally, to meet the objectives of the Regulation. As noted[336], these measures will only be effective with respect to default settings if users take advantage of these opportunities for choice and if the choice is not too burdensome for them. However, choice generally represents a significant mental burden for users, whereas interfaces are designed to be "seamless" and to make the use of the service as easy and as fluid as possible, which is particularly the case for voice assistants whose purpose is per se designed for this purpose. Indeed, users generally tend to confirm the default or most obvious choices (salience bias) or choose the names they know, or keep an existing service (status quo bias). The choice architecture is therefore decisive, which explains why the DMA prohibits deceptive choice architectures in its Article 13 (see *above*).

Moreover, it has been highlighted that so as not to make choices too complex, which supposes an adequate design and particularly: offering a clearly identified and easily accessible choice grid; devoting a section to default settings via a list of possible choices, without it being for users to choose from a list that is designed in such a way as to discourage them due to its length or complexity (see above for services that could be of interest to users, such as podcast services or playlists already installed, which they can uninstall if they do not wish to keep them); proposing a method of choice that is easy to exercise and in particular providing a command that

---

[334] Classification proposed by A. Fletcher, *DMA Switching tools and choice screens*, Issue Paper, CERRE, November 2022, p.7 and following.
[335] ARCEP, Report supra, p. 46 and following.
[336] A. Fletcher, CERRE Report supra, p. 9 and following.

makes it possible to "*make this service my default*"[337]. In this respect, it is worth referring to the CNIL's reflections on interface design by privacy, which could provide a model to follow[338].

Modalities of choice and parallelism of form. Particular attention should also be paid to modalities of choice, i.e. how choice is offered to users and how easy it is for them to exercise their right to choice. This is tantamount to imposing a principle of parallelism of form between the services of the voice assistant provider and third-party services. For example, it should be just as easy to install a music streaming application offered by the company providing the voice assistant as one offered by a third party. In other words, the latter should not be subject to an artificial complexity of installation that could discourage users (who mostly seek ease of use) from favouring it. Otherwise, it could be considered that the rule of effectiveness imposed by Article 8 of the DMA is not being respected and that this practice meets the conditions of a deceptive interface, which is prohibited by Article 13[339]. In addition, these modalities should be organised in such a way as to allow users to make an itemised choice, in order to avoid any illusion of choice that would manifest itself as a "wholesale" acceptance as opposed to a "check box" effect.

Right to setting configuration (blank). The DMA builds in a right to setting configuration[340] by lifting all setting locks, through the ability to uninstall applications, search engines, or browsers (or making the voice assistant "blank"), to choose default values (e.g., a music service application that will be the default upon user request), to go directly through third-party companies. This seems essential when it comes to voice assistants, given the seamless nature of the interface as well as the mechanism of the single response.

c. Freedom of choice conditioned by the user's mobility

Raising the barriers to exiting. Finally, user freedom of choice is conditioned on the ability to switch from one service to another without undue constraint. The reforms of the European Union law reflect this objective to remove barriers to exiting. First of all, it will be easier to unsubscribe from services. To this end, Article 6.6 of the DMA will require gatekeepers to allow users to switch between services without technical or

---

[337] A. Fletcher, CERRE Report supra, p. 16.

[338] V. CNIL, *La forme des choix. Données personnelles, design et frictions désirables*, Cahiers IP No. 6, 2019, p. 38 and following and more recently work carried out within the framework of the "Data and Design" workshop.

[339] On this notion of parallelism, see the decisions of the CNIL concerning modalities for refusing cookies: Deliberation 17 September 2020 No. 2020-092 and Deliberation 31 December 2021, No. 2021-023 and No. 2021-024.

[340] Regarding this right to setting configuration for digital services, see CNCDH, *Avis sur la lutte contre la haine en ligne*, July 2021, No. 3.2.2 : https://www.cncdh.fr/publications/avis-sur-la-lutte-contre-la-haine-en-ligne-2021-9. See also, Conseil d'Etat, *Les réseaux sociaux : enjeux et opportunités pour la puissance publique*, Annual Study 2022, No. 3.1.2 : file:///Users/celiazolynski/Downloads/EA22_r%C3%A9seaux%20sociaux_INTERNET.pdf.

other restrictions, and to subscribe to different services[341]. Making this mobility possible will also mean strengthening the effectiveness of the right to data portability.

Towards more effective data portability. The right to portability of personal data has already been enshrined in Article 20 of the GDPR; data subjects can now migrate from one operator to another, without having to give up their data or having to go through a tedious recovery process. Several years after the text came into force, its effectiveness remains very limited, given the conditions for exercising this prerogative resulting from the GDPR: the right to portability only applies to data processed automatically and not enriched, on the legal basis of consent or contract performance, and on condition that it does not infringe on the rights of third parties. There are also technical reasons[342] for this, even though industry stakeholders and trade associations have been encouraged to work together to implement a common set of interoperable standards and formats to meet these requirements[343].

To remove these lock-in effects, the DMA now recognizes a new ex ante obligation that relevant gatekeepers will have to guarantee real-time portability of their data to users of their services, free of charge and if requested by the users. Article 6.9 states that, "*The gatekeeper shall provide end users and third parties authorised by an end user, at their request and free of charge, with effective portability of data provided by the end user or generated through the activity of the end user in the context of the use of the relevant core platform service, including by providing, free of charge, tools to facilitate the effective exercise of such data portability, and including by the provision of continuous and real-time access to such data*". Portability, which remains subject to the user's request, will therefore be greatly facilitated. This facilitated portability is also discussed for all connected objects in the framework of the proposed "Data Act" regulation[344]. This text, which is still being negotiated at this time, could then allow consumers to significantly control the use of their data generated by the use of these portals and foster the development of "*competitive offer of aftermarket services, as well as broader data-based innovation and the development of products or services unrelated to those initially purchased or subscribed to by the user*" [345]. Notice that the effectiveness of real-time portability may depend on the development of data sharing standards and protocols that guarantee interoperability [346].

---

[341] See also, with respect to termination, Article 6.13 of the DMA: "*The gatekeeper shall not have general conditions for terminating the provision of a core platform service that are disproportionate. The gatekeeper shall ensure that the conditions of termination can be exercised without undue difficulty*".
[342] On this point, see J. Krämer, P. Senellart, A. de Streel, *Making data portability more effective for the digital economy: economic implications and regulatory challenges*, CERRE, 2020.
[343] EDPB, *Guidelines 02/2021 on Virtual Voice Assistants*, July 2021 – See for e.g. the Data Transfer Project, supported by Google, Apple, Meta, and Twitter, among others, to develop an open platform so that users can transfer their data to any platform that has signed on.
[344] Proposal for a Regulation establishing harmonized rules for fair access to and use of data (Data Act), COM(2022) 68 final, 23 February 2022.
[345] Explanatory Memorandum to the proposed Data Act Regulation supra, p. 16.
[346] See in particular Marx Planck Institute, *Position statement for Innovation and Competition on the Data Act proposal*, May 2022, point 301.

This could be particularly useful if the text is to be applied to voice assistants, especially if we consider the large-scale deployment of conversational agents. Indeed, these new forms of human-machine interaction, based on a natural language dialogue on the model of conversation, will make it possible to collect a large amount of data--especially emotional data--and to generate empathy mechanisms that could have important consequences on human autonomy. The control of one's data and the freedom to choose one's interface will therefore be decisive for the user.

APPENDIX 1 - Presentation of the main conversational agent models

Diversity of voice assistant models. Voice assistants are intelligent software that respond to voice commands and can run on any device, including smartphones, speakers, desktops/laptops, tablets, wearable devices, game consoles, connected speakers, virtual reality headsets, cars, and IoT devices. Overall, two different categories of voice assistants can be distinguished:
- generalist assistants whose objective is to respond to the user's request with a service (listen to a song) or an answer (provide the weather for tomorrow in Paris); examples include Amazon's Alexa, Apple's Siri, Google Assistant, and Microsoft's Cortana;
- assistants that target a specific domain and are optimized for that domain; for example, Sonos Voice Control is specialized in music, while Orange's Django is specialized in telecom technical support

Topical examples. We present the most popular commercial examples of voice assistants in the world and in France (1) as well as some examples of the next generation of chatbots to highlight the research direction taken in this field and the issues it raises (2).

1- General and specialized voice assistants

Amazon Alexa
Amazon Alexa is a voice assistant developed by the company Amazon for its Echo, Echo Dot, and Echo Show devices. Alexa is used to play music, answer general questions, set alarms and timers, or control networked devices. Recent industry statistics indicate rapid adoption of this technology, as evidenced by sales figures that have grown from 2.4 million units in 2015 to 5.2 million in 2016 to 10.6 million units sold by 2021.

Specifically, Alexa is AI software designed to perform voice functions, communicating over a WiFi connection, with Amazon's cloud servers or other networked devices, to perform those functions. In addition to getting data from Amazon's servers, the software can be used to control smart home devices, such as lighting and air conditioning. Alexa is activated when its voice recognition software receives a trigger word. For example, the word "Alexa" can be used to activate the device. This trigger word can be customized by the user. Speakers that incorporate Alexa have multiple microphones that implement noise cancellation and far-field speech recognition, so they can pick up speech in any direction and in noisy conditions.

In concrete terms, users make a request for the Alexa voice assistant. This request is filtered by Alexa using voice recognition and natural language understanding. Alexa accesses web-hosted services and provides a response to the user. As part of the response process, Alexa produces a "map" of information providing a record and history of the results to the user. This information is made available to users in their Alexa application.

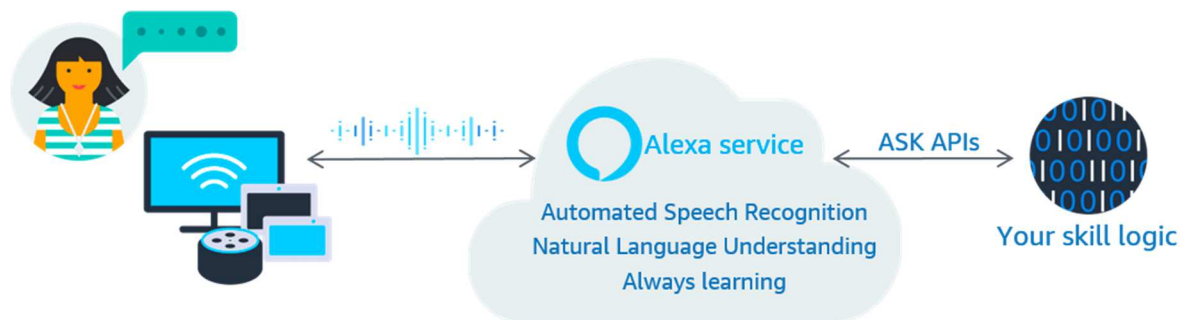
Figure -- How Amazon Alexa works [Source: What is the Alexa Skills Kit? https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html

A central aspect of Alexa are the skills. Skills are like applications for Alexa. Through its voice interface, Alexa allows users to interact with the various skills (i.e., applications installed by the user on the speaker). The user can access the content of a skill by asking Alexa to implement it. Alexa recognizes speech, determines what the user wants, and then sends a request to implement the skill that can fulfill the request. The Alexa service handles speech recognition and Natural Language Processing (NLP). The skill is however run as a service on a cloud platform. Alexa communicates with the skills registered in its application store using a request-response mechanism (through the HTTPS protocol). Technically, when a user requests an Alexa skill, it receives a POST request containing a JSON code. The body of the request contains the parameters necessary for the skill to understand the request, execute it, and then generate a response.

Each skill registered in Alexa's app store has a voice interaction template that defines the words and phrases users can speak for the skill to execute their request. This model determines how users communicate with skills. A voice user interface is similar to a graphical user interface in a traditional application, but instead of clicking buttons and selecting options, users make requests and answer questions by voice. When the user asks questions and makes requests, Alexa uses its interaction model to interpret the request and makes a specific request to a skill automatically identified always by Alexa.

In particular, Alexa offers two different voice interaction models:
1. A prebuilt speech interaction model where the ASK query defines the set of words that users must speak to invoke a skill.

2. Customized voice interaction model where skill owners have to design the voice interaction model themselves by defining all the ways a user could communicate the same request to the skill.

A skill is therefore an application capable of accepting Alexa's requests and returning the appropriate answers. During development, developers can test their skills using an Alexa simulator. Before a skill can be published in the Amazon Alexa Skills Store, Amazon must certify the skill to ensure it meets quality, security, and policy guidelines[347].

Apple Siri

Siri is the voice assistant proposed and marketed by Apple. Like Amazon Alexa, it performs searches and executes actions in response to voice commands from the user. Users of any Apple device (smartphone, tablet, computer, Apple TV, iPod touch, Apple Watch or HomePod audio speaker) can access Siri features using an Apple ID.

As of iOS 14.5 (April 2021), Siri no longer has a default voice. Users can choose the voice that speaks to them when they first set up their device, and in English, users can select more diverse voice options. This automatic speech generation is based on Neural Text to Speech technology, which aims to make speech as natural as possible. The default trigger word is "Hey Siri". Apple devices are equipped with a specialized voice recognition system that is always listening for that trigger word. However, this function must be activated by users on their device by selecting the option "Listen for 'Hey Siri'". Ideally, the sensor would trigger whenever the user says "Hey Siri", and would not trigger at other times. However, two types of errors are possible: triggering at the wrong time and not triggering at the right time. The false alarm rate is the number of false activations per hour, and the false rejection rate is the proportion of activation attempts that fail. For a given detector model, the balance between the two types of error can be modified by changing the activation threshold.

During the development of the Siri voice assistant, it was necessary to estimate the accuracy of the system using a large set of tests. There were "positive" and "negative" data. The "positive" data contained the target sentence. It was not possible to use the utterances captured by the "Hey Siri" system since the system did not capture failed attempts, and to improve the system it was important to include as many of these failed attempts as possible. Initially, the "Hey Siri" statements that some users uttered while pressing the Home button, without the intention of triggering Siri were used. These examples were not enough because the goal was for "Hey Siri" to work across a room. For this purpose, recordings were made under different conditions, such as in the kitchen (close up and at a distance), the car, the bedroom, and restaurants, by native speakers of each language. These "negative" data were used to test for false

---

[347] https://developer.amazon.com/en-US/docs/alexa/custom-skills/certification-requirements-for-custom-skills.html

activations (and false awakenings). The data represents thousands of hours of recordings, from a variety of sources, including podcasts and non-"Hey Siri" inputs to Siri in many languages, to represent both background sounds (particularly speech) and the types of phrases a user might say to another person.

When a partner wants to define an app and wants to use Siri as the interface for that app, they need to use SiriKit to define the tasks and actions that their app supports. This allows users to use Siri to perform these actions even when the app is not running. Siri can be asked to perform a system-defined task supported by the app, such as sending a message, playing a song, or starting a workout. When a task is available in your app via Siri, it is possible to customize the user experience. To reinforce this functionality throughout the user experience, the app's partner owner can write dialogues that reflect the style and tone of the company's communications and design a custom user interface that integrates the app's visual style into the Siri interface.

First, the app and SiriKit need to agree on the meaning of the request in the resolution phase. This phase represents the moment when the user is asked everything the application needs and, if necessary, requests additional information or clarifications. For example, if a person asks to send a message to Julie and they have multiple contacts named Julie, a messaging application can ask Siri to ask which Julie it is. Details about an intent, such as in this example the name of the recipient of a message, are called parameters. In the resolution phase, the partner who owns the application indicates which parameters are necessary to perform an action and which are optional. The second phase--known as the confirmation phase--occurs when all the information needed to process the user's expressed intention is available. This phase can give users the opportunity to ensure that they really want to accomplish the task. For example, tasks that have a financial impact require confirmation. Third, the application performs the task and tells SiriKit what it did during the processing phase. It is possible to provide visual and textual information that tells users what the application has done to process their request. For example, an application that allows users to order coffee may present a receipt that describes the order. Siri can speak or display the information provided by the application.

Siri can also use metadata such as location, time of day, and type of movement (such as walking, running, or driving) to automatically predict the right time and place to suggest actions from within the app. Based on the information shared by the app and the user's current context, Siri can offer shortcut suggestions on the lock screen, in search results, or on the Siri watch face. Siri can also use certain types of information to suggest actions supported by system applications, such as using Calendar to add an event shared by the application. For example, after the user uses a box office app to purchase movie tickets, Siri could remind the user to turn on the "Do Not Disturb" option shortly before the show starts.

Siri performs natural language processing for the app, allowing users to interact with the app in a variety of conversational ways. Personalized intents provide additional

opportunities to personalize the conversational dialogue, but also require users to create and speak a specific phrase to trigger the interaction. Even if the owner of an application does not write a custom dialog for the intentions provided by the system, it is possible to enhance the voice experience in other ways. For example, if the user asks Siri to play a certain music app when they enter their gym, the app can respond by playing their workout playlist.

Google Assistant

Google Assistant is a voice-activated conversational virtual assistant created by Google that can perform actions for the user and provide contextual information. Google Assistant was first unveiled at the 2016 Google I/O developer conference. Google Assistant is available on some high-end Android phones, the iPhone, Google Home, Allo (Google's smart chat app), Android TV, some smartwatches and Wear OS devices, and some cars with Android Auto integration.

Google Assistant can provide contextual information and perform actions such as booking a restaurant or sending a message on behalf of the user. Users can also type requests to Google Assistant if they don't want to use voice input. As with Alexa and Siri, Google Assistant relies on natural language processing technologies to understand what the user is saying and make suggestions or take action based on that linguistic input. Google Assistant can recognize questions or commands from multiple users and offer hands-free phone calls. It can also push the visual responses of a received request to the user's smartphone. Users trigger Google Assistant listening by saying "Hey Google" or "OK Google". In the past, users had to say one of these phrases every time they wanted to trigger the assistant, but a new feature called Continued Conversation, unveiled in 2018, allows for back-and-forth conversations, as Google Assistant can spot clues in the conversation to continue listening for commands.

While it could initially perform only one task, Google Assistant now comes with a new feature called Routines that allows the assistant to manage a chain of multiple actions with a single voice command. An example of this would be saying "Hey Google, I'm home", which triggers the lights to come on and music to start playing. As with Alexa with skills and Siri with applications, it is possible to link for example music services and choose a default music service on Google Assistant. The default service (e.g., YouTube) will be used first if possible (depending on the understanding of the request by the voice assistant).

Microsoft Cortana

Microsoft Cortana is a voice assistant marketed by Windows, which made its first appearance in 2014. It is available as an app from Windows 10 and works in conjunction with Microsoft 365 productivity software to help users to complete

business tasks. Its name is Cortana, named after a female artificial intelligence character in Microsoft's popular Halo video game series. Cortana was directly integrated into the Windows search box; the user could search on their computer or on the web by speaking or typing, and the results were delivered in a Cortana window. In addition to searching, it is possible to ask Cortana factual questions or perform tasks such as adding appointments to the calendar, sending reminders, creating lists, opening applications on the computer, checking the weather, and viewing the latest news.

In February 2020, Microsoft announced that a wide variety of consumer "skills", including music and home control, were being removed from Cortana, with the assistant being repositioned as "a personal productivity assistant in Microsoft 365". During the Covid-19 pandemic, the quality of information provided by Cortana when asked a question was problematic. An investigation by Computerworld in August 2021 found that a year and a half into the pandemic, Cortana was not providing even the most basic health information about COVID-19 and discrediting misinformation about the virus, responding "I'm sorry but I can't help you with that" or a similar non-response.

Djingo (Orange)

In France, Orange has also developed its own virtual assistant for professionals since October 2017, the virtual assistant named Djingo. This virtual assistant is designed to perform tasks such as testing fiber connection eligibility, checking bills, or setting up call forwarding, as well as to provide technical assistance in case of difficulties on the business user's mobile device.

If professional users need information, they have to type in a few words and Djingo will either answer or direct them to a relevant department. The spectrum of functionality of this virtual assistant includes SIM card activation, billing information, general practice information, Internet and landline technical support, as well as routing to a consultant or redirection to other digital departments.

The automatic natural language processing methods employed by the virtual assistant work well on grammatically correct texts. In the case of Django, the performance remains good because texts by users of this bot are generally well written. Interactions with the chatbot are generally brief, with an average of 11.3 exchanges. This length is of course variable depending on the moment when the user suspends the interaction, and the chatbot's ability to satisfy the user.

The conversation is very guided in its progression, and it is structured into three stages: the expression of the user's need, the response to this need, and the user's feedback on the progression of the interaction. An essential part of the virtual assistant is dedicated to the correct interpretation of the first request expressed by the user, and to its translation into a formalized need, a series of listed solutions. Technically, this

mechanism is presented in the form of a list of "intentions" to which the bot is able to respond, and of "variations", a list of user formulations that can correspond to these intentions. Orange is conducting an ongoing effort to review interaction logs to identify these speech patterns in order to improve the effectiveness of intent recognition. When chatbot users first express a need, they are guided by the script for the intention that has been recognized. In this phase, the interaction makes as little use as possible of the user's free expression, but instead proposes limited responses in the form of buttons. This three-act sequence of the exchange with the virtual assistant results in a strong prevalence of button responses in the interactions. If the user's sentences are too long, Django will propose a simplified rephrasing or multiple-choice answers. When the expressions are too short for the virtual assistant to decide between several answers, Django suggests that users formulate a complete sentence.

A number of virtual assistants specialized on a specific domain have also been developed in the last few years. For example, DoNotPay is a "robot lawyer" that helps people dispute incorrect parking tickets, apply for emergency housing, or apply for asylum if they are refugees.

Sonos Voice Control
Sonos Voice Control is a voice assistant designed specifically for music, with additional privacy features. Unlike Alexa or Google Assistant, this voice assistant does not upload any voice recordings to the cloud, but processes everything on the device. Sonos Voice Control works with Apple Music, Amazon Music, Pandora, Deezer, and the company's own Sonos Radio. Spotify and Google's YouTube Music are not yet available.

Sonos partnered with ReadSpeaker to generate the unique voice profile for its assistant, based on the voice of "Breaking Bad" actor Giancarlo Esposito. Esposito spent about 40 hours in the studio recording thousands of phrases and sentences that were then used as training data for the voice model. Esposito was also asked to read material specific to the Sonos voice assistant, even if those phrases aren't used in 1:1. Sonos uses voice recordings from its beta user community to train Sonos Voice Control. Sonos also partnered with outside contractors for additional recordings to train the speech processing algorithm. Sonos plans to continually update this data to account for new artists, oddly pronounced song names, and other borderline cases.

Sonos Voice Control defaults to the songs and artists that users have preferred in its application during previous interactions. In keeping with Sonos' interest in privacy, this voice assistant does not record user audio commands or transmit them to the cloud for processing.

2- Chatbots

Blender Bot (Meta)[348]
BlenderBot 3 (BB3) is an open-domain dialogue model deployed as an English-language conversational agent by Meta-Facebook. This conversational agent aims to converse naturally with users, who can then provide feedback to the model on how to improve its responses. BB3 inherits the attributes of its predecessors, including storing information in long-term memory and searching the Internet for information. The latest version of this conversational agent, BB3, combines conversational skills--such as personality, empathy, and knowledge--by integrating long-term memory and Internet searches to conduct meaningful conversations. Since chatbots are known to mimic and sometimes generate dangerous, biased, or offensive remarks, large-scale studies were conducted in the development of BB3; feedback was collected to improve these aspects in future chatbots.

To improve BlenderBot 3's ability to interact with users, the conversational agent was trained with a large amount of public language data. Most of the datasets used were collected by Meta, including a new dataset consisting of over 20,000 conversations with individuals and covering over 1,000 conversation topics. BlenderBot 3 has been trained to learn from conversations in order to improve the skills that users consider most important.

Technically, BB3 uses a learning algorithm called Director, which generates answers using two mechanisms: language models and ranking. Language models provide the model with the most relevant and fluent responses (based on training data), and then the ranking informs it of what is good or bad (based on human feedback evaluating the conversation). To generate a sentence, the language model and ranking mechanisms must be in agreement. Learning is performed jointly using both standard language model data and data labeled with desirable or undesirable sequences. Experiments conducted in several contexts show that the model has a learning and decoding speed competitive with standard language models while maintaining the quality of natural language generation. When the chatbot's response is not satisfactory, feedback from human reviewers is collected about it. With this data, it is possible to improve the model so that it does not repeat its errors.
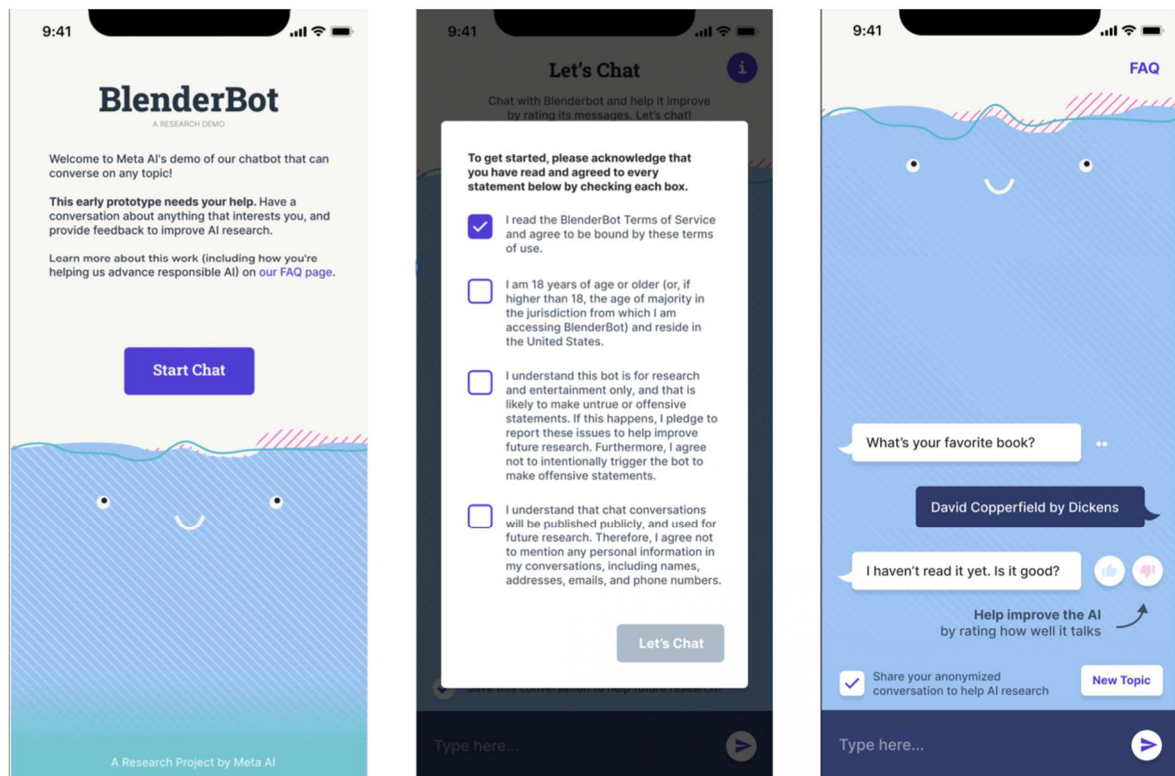
---

[348] https://arxiv.org/abs/2208.03188

Figure -- BlenderBot 3.0 user interface Source "BlenderBot 3: a deployed **conversational agent that** continually∗ learns to responsibly engage," available online https://arxiv.org/abs/2208.03188

Xiaolce (Microsoft)[349]

Xiaoice (pronounced "Shao-ice", which means "Little Bing") is a social chatbot, with a personality modeled on a teenage girl and a set of skills. In particular, in addition to the usual skills one would expect from a conversational agent, she can tell jokes, write poems, compose and sing songs, read stories, play games.

Like Samantha, the artificial intelligence played by Scarlett Johansson in the movie *Her* (2013), Microsoft's Xiaoice is as much a companion as an assistant, using empathetic computing capabilities. Since its launch in China in May 2014, Xiaoice has had over 30 billion conversations with 660 million human users worldwide. Although there are multiple ways to interact with "her", these are usually via text message. This divergence from the voice-first approach of other virtual assistants highlights a different use case. It requires a longer form of communication than just "OK Google, will it rain today?". Unlike productivity-focused assistants such as Cortana, Xiaoice is designed to have longer sessions with users. It is designed as a companion with a strong emotional component to satisfy the human need for communication, affection,

---

and social belonging. XiaoIce is optimized for long-term user engagement, measured in number of conversations per session. XiaoIce dynamically recognizes human feelings and states, understands users' intent, and responds to their needs throughout long conversations. XiaoIce's capabilities include knowledge and memory modeling, image and natural language understanding, reasoning and natural language generation. These capabilities are essential for a social chatbot to meet specific user needs and help them accomplish specific tasks. XiaoIce has developed more than 230 skills, ranging from answering questions to recommending movies or restaurants, to comforting and telling stories.

The most sophisticated skill is Core Chat, which allows for long, open conversations with users. This skill is based on two essential components: empathy and social skills. A chatbot with empathy must be able to identify the user's emotions from the conversation, detect the evolution of emotions over time, and understand the user's emotional needs. This requires query understanding, user profiling, emotion detection, sentiment recognition, and dynamic tracking of user mood in a conversation. Since users have different backgrounds, varied personal interests, and unique needs, the chatbot aims to have the ability to personalize responses that are emotionally appropriate, at times encouraging and motivating, and coherent with the user's interests. The idea is to generate socially acceptable responses (sense of humor, reassurance, etc.), and steer the conversation to a new topic when, for example, the conversation has stalled.

To train the XiaoIce chatbot, human conversations from millions of users were collected, and users were labeled as having a "desired" or "undesired" interlocutor to conduct their conversations. The majority of the "desired" users were young women. Therefore, XiaoIce was designed as an 18-year-old girl who is always reliable, friendly, affectionate, and has a sense of humor. XiaoIce tries to keep users interested by promoting a variety of conversation modes. Each conversation mode is managed by a skill that handles a specific type of conversation segment. For example, XiaoIce starts with a casual conversation using the "General Chat" skill, then moves on to a new topic about music using "Music Chat", recommends a song using the "Song-On-Demand" skill, and helps book a concert ticket using the "Ticket-Booking" skill.
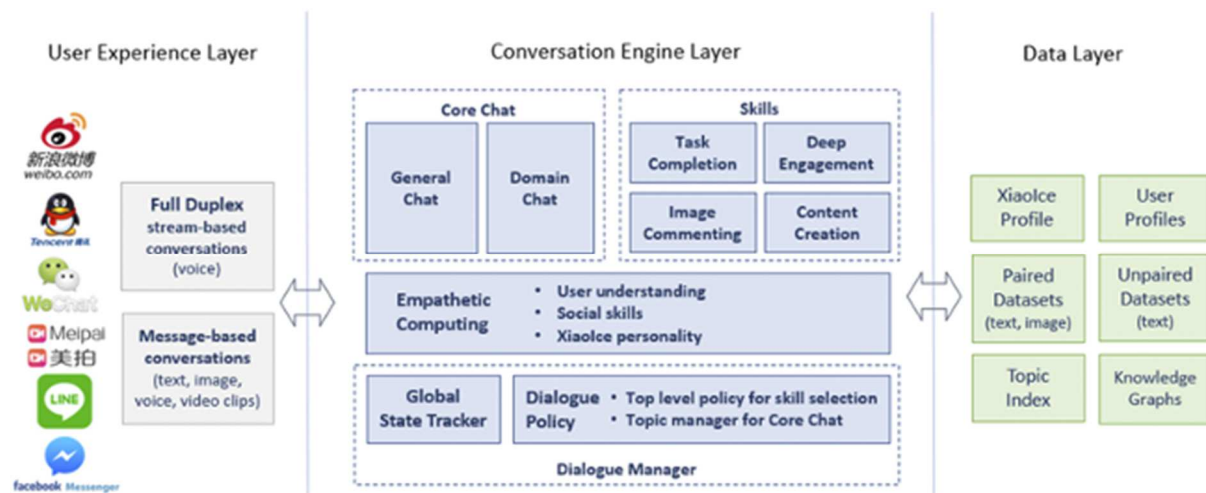
Figure -- XiaoIce System Architecture [Source "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot", available online https://direct.mit.edu/coli/article/46/1/53/93380/The-Design-and-Implementation-of-XiaoIce-and]

Replika (Luka)[350]

Replika is an interactive and personalized chatbot that learns to "reproduce" a human interaction through conversations with the user who created it. It was developed by AI startup Luka in 2017 as a way to reconnect with a deceased loved one. Replika has grown to over 10 million users worldwide, after experiencing a 35% increase during the global pandemic.

Users communicate with a Replika chatbot they have created, not with other human users. After entering their personal data, users can create their Replika. The gender (male, female, or non-binary), the appearance, and the name of the Replika are customizable. They can be changed at any time in the settings.

Replika is based on Affective computing algorithms. Replika is marketed as a friend with whom the user can talk about anything, always available at any time of the day or night.

Once the chatbot is created, users can start chatting with their "Replika". If the "live emotion" option is selected, an image of the whole body of the Replika will be visible and will "react" to the conversation. Users have access to tutorials on how to help their chatbot "become more like them" and are encouraged to use a series of feedback prompts to evaluate the responses provided by the chatbot. Conversation topics and suggested answers are encouraged to help the chatbot know what questions to ask its user. The chatbot's responses are formulated from a combination of scripted texts it has been trained on and Internet sources. Users can ask questions, play roles, flirt, take personality tests, and learn facts with their Replika.

---

[350] https://replica.ai/

Four relationship statuses are available: friend (casual chat), romantic (conversation can be sexual or intimate), mentor (goal-oriented discussions), and "see how it goes" (a mix of the previous three). The "friend" status is the only free relationship status that users can choose. This limits the scope of conversations before the user is prompted to change their relationship status.

Replika uses automatic natural language processing algorithms in all interactions. Each Replika has its own diary (in which it records what it feels about its user and its interactions) and a visible "memory" bank. This allows users to see (and edit) information stored about them, such as "Are you lonely?" or "Do you like to read books?" The chatbot adapts answers according to these "memories". "Memories" help develop the chatbot's communication skills.

APPENDIX 2 - List of people interviewed

Hearings conducted from December 2021 to November 2022

Rafael Amaro, University Professor, University of Caen

Tristan Azzi, University Professor, University of Paris 1 Panthéon-Sorbonne

Martine Behar-Touchais, University Professor, Université Paris 1 Panthéon-Sorbonne

Valérie-Laure Benabou, University Professor, UVSQ

Stéphanie Carre, Senior Lecturer, CEIPI

Jean-Marc Deltorn, Senior Lecturer, CEIPI

Joelle Farchy, University Professor, University of Paris 1 Panthéon-Sorbonne

Florence Gaullier, Lawyer

Antoine Henry, Senior Lecturer, University of Lille

Anne-Emmanuel Khan, Senior Lecturer, University of Lyon II

Florent Lafay, PhD student, University of Paris 1 Panthéon-Sorbonne

Marylou Le Roy, Post doc, University Paris 1 Panthéon-Sorbonne

Nicolas Obin, Senior Lecturer, Ircam and Sorbonne University

Adrien Pequignot, PhD student, University of Paris 8

Jean-Christophe Roda, University Professor, Lyon III University

Thibault Schrepel, Associate Professor, VU Amsterdam University

Patrice Tafforeau, University Professor, University of Lorraine

Gilles Vercken, Lawyer

ADAMI
Anne-Charlotte Jeancard
Benjamin Sauzay

Amazon
Yohan Benard
Philippine Colrat

ARCEP
Emmanuel Gabla

ARCOM
Christophe Cousin

Hervé Godechot
Benoît Loutrel
Marianne Serfaty

CNIL
Régis Chatellier
Félicien Vallet

National Digital Ethics Steering Committee
Alexei Grinbaum

European Commission (DG COMP)
Sophie de Vinck
Patricia Sanchez-Calero

National Digital Council
Jean Cattan

Deezer
Aurélien Eyrault
Nicolas Pinoteau
Mr. Ludovic Pouilly

GESTE
Sara Bahhar
Laure de Lataillade
Xavier Filliol

Meta
Antoine Bordes
Elisa Borry-Estrade
Martin Signoux

Ministry of Culture
   - **Directorate General of Media and Cultural Industries**
Mathieu Couranjou
   - **General Secretariat**
Sarah Jacquier

Ministry of Economy and Finance, Director General of Companies
Elisa Amadieu
Thierry Correard
Arthur Dinhof
Laura Hiel

Chantal Rubin

Radio France
Camille Auvret
Elsa Comby
Justine Gheeraert
Jean-Michel Orion
Fabienne Saxemard

Sonos
Joseph Dureau

Spotify
Harry Clarke
Olivia Regnie
Francesco Versace

French National Publishing Union
Laurence Ballet (Dalloz)
Léa Bernard (Madrigall)
Julien Chouraqui
Lore Vialle-Touraille (Hachette)

Tracking exposed
Marc Faddoul