



MINISTÈRE DE LA CULTURE

*Liberté
Égalité
Fraternité*

CONSEIL SUPÉRIEUR DE LA PROPRIÉTÉ LITTÉRAIRE ET ARTISTIQUE

RAPPORT DE MISSION

LES ASSISTANTS VOCAUX ET AUTRES AGENTS CONVERSATIONNELS

Présidente de la mission :

Célia Zolynski

Personnalité qualifiée au CSPLA

Université Paris 1 Panthéon-Sorbonne

En collaboration avec :

Karine Favro, Université de Haute-Alsace

Serena Villata, Université Côte d'Azur, CNRS, Inria

Décembre 2022 – Version finale

Le contenu de ce rapport, fruit d'une réflexion menée à partir d'auditions, de contributions écrites, de contacts informels et de nombreuses lectures, n'engage que ses auteurs.

SYNTHÈSE

Applications logicielles offrant des capacités de dialogue oral avec les utilisateurs, les assistants vocaux connaissent une croissance exponentielle depuis leur entrée sur le marché. S'ils ne constituent pas (encore) un point d'accès essentiel pour le secteur culturel, il s'agit particulièrement en ce domaine d'une application en forte expansion. Les assistants vocaux représentent désormais l'interface qui se développe pour permettre aux utilisateurs d'accéder au web, aux appareils connectés et aux services IoT d'une façon plus naturelle, notamment grâce aux progrès récents des méthodes d'apprentissage automatique et de traitement automatique du langage naturel (TALN). Cela s'explique en partie par le fait que la commande vocale offre de nombreux avantages dans la construction de l'interaction humain-machine, faisant de la voix une nouvelle interface, simplifiant les interactions et favorisant l'engagement de l'utilisateur. Dans le même temps, la voix comprend de nouveaux enjeux (interprétation correcte de la requête utilisateur, référencement des offres proposées, accès à l'information pour l'utilisateur, traitement de données les plus intimes) pouvant réduire le choix des utilisateurs. Et ces enjeux sont amplifiés s'agissant des assistants vocaux compte tenu du mécanisme dit de réponse unique, au risque de réduire la diversité des offres culturelles auxquelles l'utilisateur pourrait accéder via cette interface. Il en résulte également d'importants enjeux de concurrence notamment en raison de la constitution de grandes bases de données par quelques acteurs du marché et des interrogations relatives à l'accès aux données d'usage collectées, ou encore de la possibilité limitée de paramétrage et de la présence d'applications intégrées par défaut dans l'univers de l'assistant vocal.

Ces nouvelles formes d'interaction par le langage naturel supposent par conséquent de renouveler la réflexion sur la qualification et le régime juridique de ces systèmes et des effets induits sur la diffusion et la diversité des contenus culturels. A cet égard, les pratiques actuelles invitent à s'intéresser pour l'essentiel aux assistants vocaux. Il convient toutefois d'interroger aussi l'évolution de ces technologies en ce qu'elle annonce de nouvelles interfaces humain-machine suscitant un engagement accru de l'utilisateur.

A cette fin, le présent rapport, conçu dans le cadre d'une approche interdisciplinaire, propose une explicitation du fonctionnement technique des assistants vocaux et, plus généralement, des agents conversationnels ainsi que de leur univers. Cette explicitation est croisée avec les questionnements juridiques sous-jacents s'agissant du secteur culturel (I). Il envisage en outre, plus spécifiquement, les enjeux d'accès au marché et les réponses apportées par les textes récemment adoptés (II) pour ensuite analyser les conditions propres à garantir la diversité, ce qui suppose de définir le rôle que doit jouer l'utilisateur (III).

En tant que solutions technologiques, les assistants vocaux peuvent être envisagés de deux points de vue différents, mais néanmoins complémentaires, pour avoir un panorama de leur impact et des enjeux qui en dérivent. Tout d'abord, l'assistant vocal peut être étudié comme système algorithmique qui permet de traiter une requête formulée par la voix humaine et d'y répondre par une voix de synthèse. La complexité de cette architecture est liée à la collecte massive de données très hétérogènes qui, en conjonction avec des algorithmes d'apprentissage automatique de plus en plus puissants, permet de verrouiller la relation entre l'humain et l'assistant vocal à l'exclusion des entreprises utilisatrices tierces, à l'image des développeurs d'application et des éditeurs de contenus. Ensuite, l'assistant vocal doit être envisagé comme une interface permettant d'accéder à de nombreux contenus et services connexes, proposés soit par le fournisseur d'assistant vocal soit par des entreprises tierces. Il constitue en outre une fonctionnalité supplémentaire d'accès à un écosystème voire peut devenir le point d'entrée permettant d'accéder à un ensemble de services numériques proposés par son fournisseur, caractérisant une nouvelle forme d'intermédiation qui se trouve au cœur de récentes réformes du droit de l'Union européenne s'agissant des marchés et services numériques.

Concernant l'accès au marché, l'effet réseau est particulièrement développé sur le marché des assistants vocaux en ce qu'il influence les négociations entre les différents acteurs d'un marché déployé en écosystème. Les acteurs structurants occupent des positions stratégiques en la forme d'oligopole puissant mais surtout, en capacité de faire adhérer les entreprises utilisatrices à leur environnement. Ces acteurs disposent d'une capacité de filtrer les contenus, ce qui tient à la nature de moteur de résultat de l'assistant vocal. Cela contribue à renforcer considérablement le pouvoir de négociation des distributeurs de services (comme les fabricants d'enceintes connectées et autres objets connectés permettant d'étendre l'écosystème ou les fournisseurs d'assistants vocaux). De même, les fabricants de solutions techniques peuvent détenir un pouvoir de négociation significatif notamment à l'égard des éditeurs de contenus. On leur impose le coût de développement des applications vocales, le coût de l'accès au catalogue et l'effort technique consenti pour intégrer le système, tout en faisant la promesse d'une collaboration mutuelle et efficace. Enfin, le cumul d'intermédiaires puissants entre les éditeurs de contenus et les utilisateurs soulève des questions de captation et de répartition de la valeur à l'égard des éditeurs qui se trouvent en bout de chaîne contractuelle. Et l'utilisateur n'est pas préservé. Il a accès prioritairement à des services et produits mis en avant par la marque qui permet d'identifier fortement l'assistant vocal. Il est peu voire mal informé, parfois soumis à des interfaces trompeuses, des applications préinstallées par défaut, une absence de visibilité de certains contenus, ce qui peut nuire à son accès à la culture et l'enfermer dans ses choix. La réponse unique emporte une véritable restriction de l'exposition des œuvres et des services, voire de l'accès au marché.

Afin de saisir l'ensemble des enjeux s'appliquant au marché des assistants vocaux, sont examinés plus spécifiquement les points névralgiques qui permettront un accès

non discriminatoire et équitable au marché de manière à garantir la liberté de choix de l'utilisateur, notamment lors de son accès aux contenus culturels : l'autopréférence, l'interopérabilité des systèmes et des applications et l'accès aux données techniques et celles des utilisateurs. Le Règlement sur les marchés numériques (DMA) s'est saisi de ces enjeux par son approche transversale qui prévoit à cet effet des familles d'obligations pesant sur les contrôleurs d'accès.

Ces nouvelles formes d'interaction humain-machine emportent aussi d'importantes conséquences sur la diffusion et la diversité des contenus culturels. Cela suppose de penser les instruments afin de garantir la diversité au sein de ces écosystèmes. Afin que l'utilisateur exerce une liberté de choix, il convient alors d'interroger les pratiques amplifiées ou propres aux assistants vocaux dès lors qu'elles viennent altérer l'accès aux contenus pour l'utilisateur, particulièrement en raison du mécanisme de la réponse unique et de ses risques d'enfermement. Le fonctionnement de ces appareils repose également sur une collecte importante de données personnelles, par le simple enregistrement des requêtes vocales comme par l'analyse du contenu de ces requêtes qui questionne la liberté de choix de l'utilisateur.

Cela suppose d'agir sur différents leviers, selon une démarche en trois temps. De prime abord, il faut lever les restrictions qui pèsent sur le marché en interrogeant les pratiques d'autopréférence des opérateurs d'assistants vocaux privilégiant leurs propres services, notamment lorsque l'assistant vocal répond à la requête vocale de l'utilisateur en mettant en avant ses services connexes. L'autopréférence peut alors être saisie sous l'angle des pratiques restrictives d'accès au marché et des pratiques anticoncurrentielles. Au-delà de l'accès garanti à une pluralité de contenus, la liberté de choix nécessite en outre d'assurer la diversité de ces contenus. Cela conduit notamment à interroger l'intégration technique des applications existantes pour garantir la diversité, permettant non seulement de garantir l'accès à plusieurs contenus, ce que comprend le pluralisme, mais aussi de garantir l'accès à des contenus différents. Cette approche de la diversité par l'offre doit par ailleurs être complétée par une approche de la diversité par la demande. Il convient alors de penser le rôle de l'utilisateur autrement, en lui conférant le moyen d'être informé mais aussi d'agir sur l'accès au contenu, notamment en lui conférant un droit au paramétrage pour élargir son choix de contenus dans l'écosystème, mais également de choisir cet écosystème en levant les barrières à la sortie, ce qui suppose de faciliter ses conditions de désabonnement au service et de lui garantir un droit à la portabilité de ses données. Cela revient alors à penser les mécanismes de co-régulation de la diversité par l'offre et la demande.

Table des matières

Introduction	7
I - DESCRIPTION DES ASSISTANTS VOCAUX	16
A - L'assistant vocal comme système algorithmique	16
1. La requête utilisateur (Speech to Text + TAL)	24
a. Activation et écoute	24
b. Compréhension de la requête utilisateur	33
2. La génération de la réponse (TAL + Text to Speech)	37
a. La génération de la réponse (TAL)	37
b. L'expression de la réponse : la synthèse vocale (Text to Speech)	40
c. Le contenu de la réponse vocale : l'enjeu de la communication sous forme sonore d'une œuvre	50
B - L'assistant vocal comme nouvelle interface	52
1. Description de l'univers de l'assistant vocal	52
2. L'assistant vocal saisi par les textes comme une nouvelle forme d'intermédiation	56
II - ASSISTANTS VOCAUX ET ACCÈS AU MARCHÉ	65
A - L'autopréférence (ne pas bloquer l'accès au marché)	72
B - L'interopérabilité (pour accéder au marché)	77
C - L'accès aux données (pour ouvrir le marché)	85
1. La massification des bases de données des acteurs structurants	85
2. Les données pertinentes	87
3. La portabilité des données	90
III - ASSISTANTS VOCAUX ET ACCÈS AUX CONTENUS CULTURELS	92
A - Lutter contre l'autopréférence par l'expression du pluralisme : premier levier pour garantir la liberté de choix	93
1. Pluralité d'acteurs sur le marché	96
2. Diversité de modalités d'accès aux contenus	99
3. Pluralité des contenus	103
B - L'enjeu de la découvrabilité : deuxième levier pour envisager la diversité autrement	110
C - L'utilisateur : nouvel acteur du pluralisme et de la diversité	117
1. La mise en pouvoir d'agir de l'utilisateur par la donnée	119
2. Liberté de choix et interaction avec le service numérique	122
a. La liberté de choix conditionnée par l'information de l'utilisateur	122

b. La liberté de choix conditionnée par la possible action de l'utilisateur sur le système	132
c. La liberté de choix conditionnée par la mobilité de l'utilisateur	135
ANNEXE 1 - Présentations des principaux modèles d'agents conversationnels	137
1- Assistants vocaux généralistes et spécialisés	137
2- Chatbots	145
ANNEXE 2 - Liste des personnes auditionnées	150

Introduction

Présentation générale. L'année 2018 est une année marquante pour les assistants vocaux, laquelle signe une fin, celle des smartphones tels qu'ils sont conçus, et un début, celui de l'embarquement des objets connectés¹. Les utilisateurs sont saturés d'écrans qu'il s'agisse de smartphones ou d'ordinateurs, alimentés par des discours sur les enjeux de santé publique. Particulièrement depuis la crise sanitaire, cela produit un engagement pour d'autres modes de consommation et d'interaction au profit d'un regain d'intérêt pour le canal audio² et la montée en puissance des enceintes connectées intégrant généralement l'assistant vocal de Google et celui d'Amazon³. Les assistants vocaux grand public ont d'abord été accessibles sur les smartphones et autres tablettes tactiles, mais ils se développent, à cette date, sur des environnements connectés pour construire la maison intelligente de demain. L'assistant deviendra alors « *une télécommande universelle* », qui pourrait disqualifier « *les équipements qui ne sont pas compatibles avec sa couche logicielle* »⁴. L'objectif de ces prochaines années sera donc de concevoir des interfaces invisibles portées par les utilisateurs afin de les accompagner dans leurs mobilités, possiblement jusqu'aux univers virtuels. A tel point que le seuil de saturation lié au temps passé sur les écrans pourrait rapidement être dépassé avec l'écoute de nouveaux supports conçus pour divertir, puis assister l'utilisateur au quotidien⁵.

Depuis leur entrée sur le marché, la croissance des assistants vocaux est exponentielle. Désormais, « *les assistants vocaux représentent l'interface qui se développe le plus rapidement pour que les utilisateurs puissent accéder au web, utiliser et contrôler des appareils intelligents et accéder à des services IoT grand public* »⁶. Leur utilisation se veut aujourd'hui plus ambitieuse. Les progrès récents des méthodes d'apprentissage automatique et de traitement automatique du langage naturel (TALN), combinés à la disponibilité de la puissance de calcul, ont mis à disposition des développeurs de nouveaux algorithmes créés pour rendre l'interaction plus naturelle. Aujourd'hui, les agents conversationnels peuvent être considérés comme une « *technologie mature et déployée, mais elle reste perfectible* »⁷, dans l'attente d'une nouvelle génération.

Assistants vocaux, chatbot, agents conversationnels : une terminologie et des techniques plurielles. Chatbot, assistant vocal, assistant virtuel, agent conversationnel sont autant de termes employés pour désigner de prime abord la même chose, à savoir le fait de dialoguer avec une machine. Cela signifie que la

¹ HADOPI/CSA, *Assistants vocaux et enceintes connectées. L'impact de la voix sur l'offre et les usages culturels et médias*, mai 2019, p. 6.

² O. Gouliáeva, E. Dosquet, Y. Moysan, *La révolution des assistants vocaux*, DUNOD, 2020, p.11 et s.

³ HADOPI/CSA, Etude préc.

⁴ France Stratégie, *Le monde de l'internet des objets : des dynamiques à maîtriser*, février 2022, p. 107.

⁵ O. Gouliáeva, É. Dosquet, Y. Moysan, préc.

⁶ Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022, point 28 - Sur l'évolution du marché, v. HADOPI/CSA, Etude préc.

⁷ France Stratégie, Rapport préc., p. 87.

traduction du terme « chatbot » exprime clairement une action qui caractérise une interaction humain/machine présente lorsque l'on fait référence à l'assistant vocal, virtuel ou même l'agent conversationnel. S'agit-il alors de termes génériques employés indistinctement par les concepteurs et les autorités qui se saisissent de cette innovation ? La question mérite d'être posée car la réponse ne semble pas figée.

Si d'aucuns le considèrent, la réalité semble plus nuancée car la terminologie dépend du niveau de conversation. Autrement dit, le critère qui pourrait distinguer l'assistant vocal d'un chatbot procède de la construction d'un échange permettant à l'utilisateur de recevoir une réponse non seulement aux questions les plus évidentes qu'il va poser au chatbot, mais encore aux questions sous-jacentes qu'il se pose. Le chatbot peut ainsi analyser les intentions de l'utilisateur à partir d'une action simple comme la consultation des comptes bancaires, et peut même l'amener à comprendre ses intentions qu'il n'a pas toujours prémédité, comme la raison pour laquelle il consulte ses comptes bancaires. Le procédé est identique avec la météo, voire la musique, etc. Lorsque l'agent conversationnel est capable de ce niveau de conversation et de maintenir la conversation, on parle de chatbot. Le chatbot peut notamment aller au-delà en renforçant la personnalisation de la conversation et l'accompagnement quotidien de l'utilisateur. Non seulement les réponses sont personnalisées mais elles ont vocation à se fondre au comportement de l'utilisateur au quotidien. L'approche est plus contextuelle, requérant des données de contexte et comportementales. Dès lors, on passe volontiers du simple assistant (proposant une réponse du type FAQ / chat pour discussion en ligne) en ce qu'il fournit une réponse contextualisée en facilitant un service rendu à l'utilisateur, au chatbot qui est un « *majordome numérique à notre service, à notre écoute* »⁸ en ce qu'il intègre des réponses personnalisées à des demandes complexes et des processus automatisés pour exécuter certaines actions.

En réalité, l'assistant (vocal) virtuel représente la première génération d'agent conversationnel montrant une interaction humain/machine encore embryonnaire, voire limitée techniquement (par exemple, par l'interface vocale, l'interopérabilité, les conditions d'accès au marché) dans la mise à disposition des contenus ce qui interroge sur leur capacité à informer ou divertir à l'aune du respect du pluralisme. Le chatbot peut caractériser également une interaction vocale avec l'utilisateur mais elle pourrait être même plus poussée, allant jusqu'à assurer le clonage de la voix de l'utilisateur en synthèse vocale. Par conséquent, il s'agit de délivrer au chatbot des clefs de compréhension des émotions humaines ce qui constitue d'emblée un risque d'enfermement de l'utilisateur dans son propre univers.

Assistants vocaux, chatbot, agents conversationnels : définitions. Il convient dès lors d'apporter des précisions terminologiques et de présenter les définitions à retenir pour analyser les enjeux relatifs au déploiement des assistants vocaux et autres agents conversationnels constituant l'objet du présent rapport. En termes généraux,

⁸ FING, *Hypervox. Promesses et illusions des nouvelles interfaces humain machine*, janvier 2019, p. 9.

les agents conversationnels ou systèmes de dialogue⁹ sont des logiciels qui communiquent avec les utilisateurs en langage naturel (à travers du texte, de la parole, ou les deux) et se divisent en deux catégories. La première catégorie inclut les assistants virtuels (AV) qui utilisent la conversation avec les utilisateurs pour les aider à accomplir des tâches spécifiques. Certains assistants virtuels sont capables d'interpréter la parole humaine et de répondre par des voix synthétiques. Ces assistants virtuels sont appelés « assistants vocaux ». Les utilisateurs peuvent poser des questions à leurs assistants vocaux généralistes, contrôler les appareils domotiques et la lecture des médias par la voix, et gérer d'autres tâches de base telles qu'envoyer un courrier électronique à l'aide de commandes verbales. Siri, Alexa, Google Now/Home et Cortana constituent des exemples d'assistants vocaux généralistes. La technologie des assistants vocaux est incontournable pour les enceintes connectées, pour lesquelles il est l'unique mode d'interaction¹⁰. La deuxième catégorie inclut les chatbots qui sont des systèmes conçus pour des conversations prolongées, qui imitent les conversations non structurées ou les "chats" caractéristiques de l'interaction humain-humain, principalement à des fins de divertissement, mais aussi à des fins pratiques, comme rendre plus naturels les assistants virtuels. C'est le cas par exemple de BlenderBot de Meta ou du système Xiaolce de Microsoft. Certains chatbots ont une personnalité plus développée que d'autres ; ils sont alors dénommés chatbots sociaux.

Dans la suite du rapport, il sera fait référence aux assistants virtuels pour parler des agents conversationnels orientés vers une tâche et aux chatbots pour les systèmes capables d'aborder une conversation prolongée avec l'utilisateur, sans tâche précise à accomplir. Lorsque l'assistant virtuel interagit par la voix avec l'utilisateur, le terme « assistant vocal » sera préféré afin de le désigner.

Distinction opérée entre les assistants vocaux généralistes ou polyvalents et les assistants vocaux spécialisés. La Commission européenne s'est intéressée à la question des assistants vocaux dans le cadre de son enquête sectorielle sur l'internet des objets grand public¹¹. A ce titre, la définition proposée est la suivante : « *un logiciel à commande vocale qui peut traiter des commandes vocales et renvoyer des informations pertinentes ou exécuter certaines fonctions à la demande des utilisateurs* ». En effet, l'assistant vocal agit à la fois comme une plateforme d'intermédiation pour les applications vocales et comme interface utilisateur. Cette interface se développe désormais rapidement en raison de la facilité avec laquelle l'utilisateur peut formuler une requête avec la voix pour accéder au web, utiliser et « contrôler » des appareils intelligents. L'interaction vocale est utile dans certains environnements contraints comme la voiture, ou le domicile, en complément d'une autre activité. Les assistants vocaux qui dominent le marché européen sont polyvalents, en ce qu'ils permettent à l'utilisateur d'accéder à de multiples

⁹ D. Jurafsky, J. H. Martin, «Speech and Language Processing», *Prentice Hall*, 2021, <https://web.stanford.edu/~jurafsky/slp3/>

¹⁰ HADOPI / CSA, Etude préc.

¹¹ Commission européenne, Enquête préc.

fonctionnalités identifiées par la Commission. Ils se distinguent (ou absorbent selon les cas) d'assistants vocaux spécialisés fournis par les fournisseurs de services IoT ou des fabricants de dispositifs de domotique. Ils peuvent également être configurés comme de simples interfaces de contrôle vocal.

Les assistants polyvalents, selon la distinction opérée par la Commission¹², permettent de lire la musique et les vidéos, d'écouter la radio, des podcasts, des informations, des livres audio. Ces assistants vocaux sont associés à d'autres dispositifs intelligents pour faciliter l'écoute ou le visionnage des réponses aux requêtes formulées par l'utilisateur, tels que des écouteurs, des enceintes ou des téléviseurs. Cependant, l'assistant polyvalent peut également fournir des informations qui alimentent le quotidien des utilisateurs, comme la météo, un trajet, une recette de cuisine, des conseils de jardinage, une écoute musicale, un podcast, etc. Enfin, il peut participer à la routine quotidienne de l'utilisateur mettant à exécution son planning, comme les rappels d'événements dans le calendrier, la réservation des repas, la gestion des thermostats dans la maison, etc. L'assistant polyvalent tel que défini par la Commission européenne est intéressant en ce qu'il caractérise un véritable système, voire écosystème, au sein de l'écosystème d'Amazon (Alexa), de Samsung (Bixby), de Google (Google Assistant) et d'Apple (Siri). L'assistant vocal permet non seulement d'accéder aux services proposés par l'écosystème duquel il émane, mais également de recourir aux applications tierces disponibles de manière à fournir une réponse adaptée à la requête des utilisateurs. L'assistant vocal peut, en effet, exécuter différentes tâches telles que la captation et la restitution du son, la transcription automatique de la parole (à l'aide d'une conversion de la parole en texte : « speech to text »), ou l'inverse en générant du langage (la synthèse vocale qui permet de traduire le texte en parole : « text to speech »), un traitement automatique du langage, des stratégies de dialogue, un accès aux ontologies (des jeux de données structurées liées à un domaine défini) et sources de connaissances externes. L'assistant vocal n'est pas en soi intelligent mais est intégré à des systèmes intelligents, ce qui permet de confondre les deux.

D'autres assistants sont considérés comme spécialisés¹³ en raison de leurs fonctionnalités limitées aux services proposés par les fournisseurs de services IoT qui les mettent sur le marché. C'est le cas de l'assistant vocal d'Orange, Djingo, autorisant l'accès aux services Orange, mais également de OK Freebox, OK SFR, Cortana de Microsoft, etc. L'assistant spécialisé, limité par son environnement audiovisuel ou domotique, est généralement disponible dans une seule langue, ce qui n'est pas le cas des assistants polyvalents qui oscillent entre 8 et 30 langues¹⁴. En principe, il n'y a pas d'interaction entre les assistants vocaux généralistes et les assistants vocaux spécialisés.

¹² Commission européenne, Enquête préc., point 30.

¹³ Commission européenne, Enquête préc., point 31.

¹⁴ Commission européenne, Enquête préc., point 32.

La terminologie retenue par les textes. Les lignes directrices visant à appliquer le RGPD aux « assistants vocaux virtuels » adoptent formellement la bonne dénomination pour désigner cette forme d'interaction vocale, qui y est définie comme un « *service qui comprend les commandes vocales et les exécute ou assure leur médiation avec d'autres systèmes informatiques si nécessaire* »¹⁵. L'objectif de ces lignes directrices étant de préconiser une marche à suivre s'agissant de l'utilisation de la voix comme donnée personnelle, l'accent est mis sur la commande vocale, qui caractérise somme toute l'élément central du système. Le Règlement sur les marchés numériques (ou Digital Market Act)¹⁶ intègre dans son champ d'application la catégorie plus générique d'assistant virtuel. L'article 2 point 12) du règlement définit ainsi l'« assistant virtuel » en la forme « *d'un logiciel qui peut traiter des demandes, des tâches ou des questions, notamment celles fondées sur des données d'entrée sonores, visuelles ou écrites, de gestes ou de mouvements, et qui, sur la base de ces demandes, tâches ou questions, donne accès à d'autres services ou contrôle des appareils connectés physiques* ». La démarche portée par le texte se veut plus globale et prospective, ne se limitant pas seulement au marché de la voix, mais vise l'ensemble des questions d'accès au marché qui concerne également l'interopérabilité, l'autopréférence, les aspects contractuels, etc. Cette définition est d'ailleurs reprise par la proposition de règlement Data Act¹⁷. Ces textes saisissent ainsi l'assistant vocal/virtuel dans sa fonction d'intermédiation en ce qu'il donne accès à des services fournis par la plateforme et les entreprises utilisatrices.

Les applications vocales nécessaires au fonctionnement de l'assistant vocal.

On comprend dès lors en quoi l'étude ne saurait se limiter au seul assistant vocal mais doit appréhender l'ensemble de son « univers », et tout particulièrement les applications vocales qui sont requises pour répondre aux demandes des utilisateurs. Celles-ci sont définies par la Commission européenne comme un « *logiciel conçu pour un assistant vocal spécifique qui prend en charge les commandes des utilisateurs pour se connecter à des dispositifs smart, pour exécuter des actions ou des tâches, ou pour s'engager dans des services IoT pour consommateurs auxquels les consommateurs accèdent via cet assistant vocal* ». Leur dénomination peut être diverse : celles écrites pour Amazon Alexa sont appelées « compétences » (ou *skills*), pour Google Assistant « actions », pour Apple les « siri shortcuts ». Au-delà de leur dénomination, ces applications sont spécifiques à chaque assistant vocal et un développeur devra proposer une application vocale différente pour chacun d'entre eux. Cela doit être mis en perspective avec le modèle économique des fournisseurs d'assistants vocaux, décrit par l'enquête sur les marchés digitaux publiée par la United States House of Representatives en juillet 2022¹⁸. D'autres études dressent quant à elles le constat

¹⁵ CEPD, *Lignes directrices sur les assistants vocaux virtuels*, 02/2021, 7 juillet 2021, p. 3.

¹⁶ Règlement (UE) 2022/1925 du Parlement européen et du Conseil du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (Règlement sur les marchés numériques).

¹⁷ Proposition de Règlement du Parlement européen et du Conseil du 23 février 2022 fixant des règles harmonisées pour l'équité de l'accès aux données et de l'utilisation des données (règlement sur les données), COM(2022) 68 final, article 2.4.

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, juillet 2022, p. 100 : « A voice assistant

selon lequel les assistants vocaux sont dans les faits un « *continuum de services* » déjà existants, et « *suivent la stratégie des acteurs qui les opèrent* »¹⁹.

La voix, une nouvelle interface humain-machine. La commande vocale offre de nombreux avantages dans la construction de l'interaction humain-machine, en ce qu'elle porte un confort d'utilisation passant de la voix au tactile et vis-versa. Elle permet ensuite d'exercer plusieurs activités simultanément en ayant les mains occupées, de garantir l'immédiateté de l'interaction, tout en étant seulement à portée de voix, et parfois sans être dans le périmètre de l'assistant vocal²⁰. A cela s'ajoute « *le caractère naturel de l'interaction qui n'implique pas d'apprentissage spécifique de la part des utilisateurs, la rapidité de l'exécution de la commande et l'extension du champ d'application qui permet un accès rapide à l'information* »²¹. La voix s'imprègne comme une interface naturelle permettant d'oublier la technologie. Elle pourrait ainsi devenir le nouveau standard de l'interaction humain-machine. A ce jour, elle vient en complément d'autres interfaces graphiques (comme les écrans des ordinateurs et smartphones) mais pourrait à terme les remplacer pour devenir l'interface principale de divers appareils informatiques, notamment les appareils d'IoT.

La voix et la réponse unique. Si elle simplifie les interactions, la voix comprend dans le même temps de nouveaux enjeux qu'il s'agisse de l'interprétation correcte de la requête utilisateur, du référencement des offres proposées et plus généralement de l'accès à l'information pour les utilisateurs des assistants vocaux, notamment s'agissant du traitement de leurs données à caractère personnel massivement collectées par ce type d'interface. Elle pourrait en outre réduire le choix des utilisateurs, particulièrement en cas d'applications intégrées par défaut et de possibilité limitée de paramétrage. En outre, le fonctionnement de l'assistant vocal, et particulièrement la réponse unique, permettent d'établir un profil détaillé de l'utilisateur par l'accumulation de données ciblées par leurs requêtes. L'apprentissage automatique en temps réel est l'une des principales caractéristiques de l'assistant vocal, lequel lui permet d'apprendre sur ses habitudes pour adapter au mieux sa réponse. L'objectif avéré résulte des conséquences d'une interaction régulière pour anticiper les besoins contextuels de l'utilisateur et lui offrir une assistance préventive sans attendre ses requêtes.

Dans le secteur culturel, tout particulièrement, la création audio présente des caractéristiques propres liées à l'utilisation de la voix, le travail du son, le récit, l'environnement musical, la dimension esthétique particulière de l'écoute, à l'origine de nouveaux usages. Cela permet par ailleurs d'atteindre de nouveaux utilisateurs

platform vendor 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 janvier 2019 - pour une description de ces modèles, v. CNIL, *A votre écoute. Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*, Livre blanc n°1, 2019, pp. 24-25.

²⁰ O. Gouliáeva, É. Dosquet, Y. Moysan, *La révolution des assistants vocaux*, DUNOD, 2020, p.11 et s.

²¹ CEPD, *Lignes directrices sur les assistants vocaux virtuels*, et déjà CNIL, *A votre écoute*, préc., pp. 19-20.

parmi les plus jeunes et de faire émerger de nouveaux talents et notamment en matière de podcasts²². Au-delà, la voix facilite l'accès aux contenus culturels et permet d'affiner la relation avec l'utilisateur en collant au contexte d'écoute mais, dans le même temps, peut réduire le choix de l'utilisateur en raison de la réponse unique de l'assistant vocal.

Assistants vocaux pour le secteur culturel. Les assistants vocaux ne sont pas encore un point d'accès essentiel pour le secteur culturel mais il s'agit d'une application en forte croissance. Dans le domaine de la musique, par exemple, Sonos a récemment introduit son propre assistant vocal, le Sonos Voice Control. Celui-ci est conçu spécialement pour la musique et doté de garanties de confidentialité additionnelles. Le Sonos Voice Control utilise comme réponse par défaut les chansons et les artistes que les utilisateurs ont préférés dans son application, lors des précédentes interactions. Contrairement aux assistants vocaux généralistes tels que Amazon Alexa ou Google Assistant, cet assistant vocal ne télécharge aucun enregistrement vocal sur le *cloud*, mais traite tout sur l'appareil.

Dans le domaine de la radio, une révolution engendrée par le marché des assistants vocaux est en cours. Pour l'instant, ils sont utilisés pour accéder aux contenus tels que les podcasts, mais leur impact pourrait devenir beaucoup plus important dans les prochaines années. Dans le domaine du livre, les assistants vocaux peuvent aider les utilisateurs à acheter des livres physiques. De plus, la présence sur les magasins en ligne des principaux acteurs concernés (Google, Apple et Amazon) de livres audio et de livres numériques qui pourraient être lus par l'enceinte connectée représente un potentiel de développement de ces assistants vocaux. Plus ponctuellement, des projets ont pu se développer pour tirer parti de ces nouvelles modalités d'interaction humain-machine, à l'image de projets de lecture augmentée, accompagnée d'un assistant vocal lançant en même temps que la lecture des accompagnements sonores pour favoriser l'engagement du public²³ ou encore des contenus audio augmentés pour envisager un dialogue avec les contenus²⁴. Récemment, une nouvelle enceinte connectée appelée Merlin a été proposée sur le marché par Bayard et Radio France. Il s'agit d'une conteuse d'histoires qui arrive avec 42 titres préchargés (environ 4h d'enregistrements audio) dans sa mémoire interne de 4Go. La capacité de Merlin est de 50 heures au total. Il est possible d'ajouter d'autres titres audio parmi les 140 supplémentaires proposés via l'application Merlin (disponible pour Android et iOS). Cela demande de créer préalablement un compte utilisateur sur le site officiel hello-merlin.com, avant de pouvoir gérer la bibliothèque de contenu et la synchroniser avec l'enceinte Merlin en Wi-Fi. La sélection est variée, avec des histoires écrites par des auteurs contemporains (Véronique Olmi, Marie Desplechin, Julien Blanc-Gras,...),

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 et N. Phoyu-Yedid, octobre 2020 - dans le prolongement de ce rapport, une réflexion est désormais menée au sein de l'Observatoire du podcast, sous l'égide de l'ARCOM et du Ministère de la Culture (DGMIC), notamment pour mesurer les dynamiques du secteur.

²³ En ce sens, v. le projet lancé par Disney : <https://www.vokode.com/lassistant-vocal-disney-vous-aide-a-devenir-le-meilleur-conteur-dhistoires/>

²⁴ Sur ces différents projets, v. FING, *Hypervoix*, 2019, pp. 23-25.

des documentaires, des chansons (en français et en anglais), des contes musicaux et même de la gym. Pour ce qui concerne la qualité audio, Merlin se démarque nettement par rapport aux conteuses d'histoires concurrentes. Il ne s'agit pas, en revanche, d'un assistant virtuel ou vocal, car l'interaction avec l'enceinte passe à travers des boutons (un petit bouton d'allumage et 5 boutons disposés en cercle pour naviguer les menus) et non à travers la parole ni le texte (chat).

En dehors de ces hypothèses spécifiques, on relèvera la difficulté des acteurs du secteur culturel à développer leur propre assistant vocal, ce qui les enjoint à nouer des partenariats avec les fournisseurs d'assistants vocaux généralistes, au prix de négociations le plus souvent déséquilibrées, d'autant que ces offres sont généralement mondialisées. Pour autant, ces partenariats sont indispensables dès lors que les producteurs de contenus culturels entendent accéder à une large demande. Ils se trouvent alors contraints de supporter les coûts de déploiement d'applications vocales par chacun des assistants vocaux afin d'accéder à ce public.

Quelles perspectives pour demain ? Les développements techniques permettent d'envisager à l'avenir le déploiement de chatbot avec des interfaces humain-machine reposant sur une interaction plus naturelle à travers la parole qui permettra d'adapter encore plus finement les réponses de l'agent conversationnel à la requête de l'utilisateur. Simulant une conversation humaine, le chatbot sera en capacité d'identifier les émotions ressenties par son utilisateur et de simuler en retour des émotions - qui demeurent factices - ce, afin de générer une interaction toujours plus poussée et source d'engagement en jouant sur un phénomène d'empathie²⁵. Ce phénomène pourrait demain être encore amplifié si ces chatbot se déploient dans les univers virtuels comme semblent l'annoncer certains projets actuels²⁶. Cette forme d'interaction pourrait en outre venir au soutien de nouvelles modalités de création, comme en attestent divers projets artistiques²⁷.

Par ailleurs, une attention particulière doit être portée au développement des transformeurs qui constituent une architecture de réseau de neurones permettant d'effectuer des tâches complexes telles que la génération automatique de texte ou d'images avec des coûts d'entraînement réduits²⁸. Par exemple, GPT-3 peut, à partir d'un titre, écrire un paragraphe développant l'argument contenu dans le titre. Il peut également être utilisé pour faire de l'écriture créative ; il suffit de le solliciter avec une

²⁵ Sur ce point, v. notamment Comité national d'éthique du numérique (CNPEN), Avis n°3 *Agents conversationnels : enjeux d'éthique*, novembre 2021.

²⁶ V. par ex. le projet *Builter Bot de Meta* (v. annexe 1).

²⁷ Par exemple, v. la présentation de J. Walsh, « *Zombie Creativity : the death and resurrection of the author in the age of AI* », colloque du projet TESaCO (Académie des sciences morales et politiques) « *Talking machines : conversation and creation with Artificial Intelligence* », 8 novembre 2021 :

<https://www.academie-sciences.fr/fr/Colloques-conferences-et-debats/mythes-et-machines.html> - également le projet de Stéphanie Kipkins « *Not the only one* » : <https://www.stephaniedinkins.com/ntoo.html> ou encore celui de Lauren Mc Carthy : <https://arstechnica.com/gaming/2018/01/artist-transforms-herself-into-a-virtual-assistant-and-obey-your-commands/>

²⁸ Précisions par ailleurs que les modèles basés sur les transformateurs ont commencé à avoir un impact significatif en dehors du langage naturel. Par exemple, en vision artificielle, le transformeur de vision atteint une précision impressionnante dans les tâches de classification d'images en traitant de petits morceaux d'une image comme des éléments d'une séquence, tout comme une phrase composée par des mots.

courte phrase. Ces systèmes jouent déjà un rôle non négligeable dans le secteur de la presse écrite comme en atteste le fait que plusieurs journaux dont le New York Times, The Associated Press, Reuters, The Washington Post et Yahoo! Sports, utilisent des algorithmes d'apprentissage automatique pour générer du contenu. GPT-3 est un rédacteur d'opinion pour The Guardian et The Associated Press, par exemple, peut désormais produire 30 000 articles d'actualité locale par mois grâce à l'IA. Ces modèles sont utilisés aussi pour construire les environnements de réalité virtuelle et de réalité augmentée qui sont à la base des métavers.

Enjeux. Compte tenu de ces multiples enjeux, ces nouvelles formes d'interaction par le langage supposent de renouveler la réflexion sur la qualification et le régime juridique de ces systèmes. Le déploiement des agents conversationnels soulève des questions déjà envisagées telles que celles liées à la protection des systèmes algorithmiques²⁹. D'autres sont encore à analyser, à l'image des effets induits par ces nouvelles formes d'interaction humain-machine sur la diffusion des contenus culturels. On observe par exemple que le mécanisme dit de réponse unique, la structuration de l'interface et son écosystème peuvent réduire la diversité des offres culturelles auxquelles l'utilisateur pourrait accéder via cette interface. Il en résulte également d'importants enjeux de concurrence notamment en raison de la constitution de grandes bases de données par quelques acteurs du marché et des interrogations relatives à l'accès aux données d'usage collectées, ou encore de la possibilité limitée de paramétrage et de la présence d'applications intégrées par défaut dans l'univers de l'assistant vocal.

Champ de l'étude. Compte tenu des pratiques actuelles, l'étude portera pour l'essentiel sur les assistants vocaux afin de répondre aux enjeux qui résultent de leur déploiement croissant, tout en interrogeant l'évolution de ces technologies en ce qu'elle annonce de nouvelles interfaces humain-machine suscitant un engagement accru de l'utilisateur. Le présent rapport, conçu dans le cadre d'une approche interdisciplinaire, propose dès lors une explicitation du fonctionnement technique des assistants vocaux et, plus généralement, des agents conversationnels ainsi que de leur univers, croisée avec les questionnements juridiques sous-jacents s'agissant du secteur culturel (I). Il envisage plus spécifiquement les enjeux d'accès au marché et les réponses apportées par les textes récemment adoptés (II) pour, en outre, analyser les conditions propres à garantir la diversité, ce qui suppose de définir le rôle que doit jouer l'utilisateur (III).

²⁹ V. tout particulièrement A. Bensamoun et J. Farchy, *Intelligence artificielle et Culture*, rapport final présenté au CSPLA, janvier 2020.

I - DESCRIPTION DES ASSISTANTS VOCAUX

Les assistants vocaux sont des applications logicielles offrant des capacités de dialogue oral avec les utilisateurs. Ces solutions technologiques peuvent être envisagées de deux points de vue différents, mais néanmoins complémentaires, pour dresser un panorama de leur impact et des enjeux qui en dérivent. D'une part, l'assistant vocal peut être étudié comme système algorithmique qui permet de traiter une requête formulée par la voix humaine et d'y répondre par une voix de synthèse. C'est toute la richesse de cette technologie, car l'envisager comme un système algorithmique montre la complexité de l'architecture en amont de l'interaction entre l'assistant vocal et l'utilisateur. Cette complexité est liée à la collecte massive de données très hétérogènes qui, en conjonction avec des algorithmes d'apprentissage automatique de plus en plus puissants, permet de verrouiller la relation entre l'humain et l'assistant vocal à l'exclusion des entreprises utilisatrices tierces, à l'image des développeurs d'applications et des éditeurs de contenus (A). D'autre part, l'assistant vocal doit être envisagé comme une interface permettant d'accéder à de nombreux contenus et services faisant partie de son univers. Il constitue en outre une fonctionnalité supplémentaire d'accès à un écosystème, voire peut devenir le point d'entrée permettant d'accéder à un ensemble de services numériques proposés par son fournisseur, caractérisant une nouvelle forme d'intermédiation (B).

A - L'assistant vocal comme système algorithmique

L'assistant vocal va permettre à l'être humain de parler à un « terminal » donc à la machine, de lui adresser une requête orale, à laquelle il va répondre grâce à un système de reconnaissance vocale et de traitement du langage naturel. C'est un système algorithmique qui reçoit des données liées à l'humain ou la machine pour en déduire une manière d'atteindre une série d'objectifs en utilisant l'apprentissage, le raisonnement ou la modélisation dans la perspective de générer des résultats sous forme de contenus, de prédictions, de recommandations ou de décisions influençant l'environnement réel ou virtuel avec lequel ils interagissent. Par ailleurs, la formulation de cette réponse par une voix de synthèse représente un coût non négligeable en raison de développements complexes, et s'accompagne de possibilités d'erreurs d'interprétation bien plus importantes qu'une requête écrite. Pour autant, il répond à un confort d'utilisation plébiscité notamment dans les univers contraints pour des raisons de sécurité, tel que la voiture. Parce qu'il se commande à la voix et qu'il laisse les mains disponibles, il a vocation à s'embarquer dans le quotidien. En cela, cette nouvelle interaction humain-machine devient un vecteur d'exposition des contenus culturels dont le développement va croissant depuis la mise sur le marché des premiers assistants vocaux en 2011.

Agents conversationnels. Définition(s). Les agents conversationnels ou systèmes de dialogue³⁰ sont des logiciels qui communiquent avec les utilisateurs en langage naturel (texte, parole, ou les deux) et se divisent en deux catégories.

D'une part, les **agents de dialogue orientés tâches ou assistants virtuels** utilisent la conversation avec les utilisateurs pour les aider à accomplir des tâches. Plus précisément, un **assistant virtuel intelligent** est un logiciel capable d'exécuter des tâches ou des services pour un individu en fonction de commandes ou de questions. Certains assistants virtuels sont capables d'interpréter la parole humaine et de répondre par des voix synthétiques. Ces assistants virtuels sont appelés **assistants vocaux**. Les utilisateurs peuvent poser des questions à leurs assistants vocaux généralistes, contrôler les appareils domotiques et la lecture des médias par la voix, et gérer d'autres tâches de base telles que le courrier électronique et les calendriers à l'aide de commandes verbales. Les **assistants vocaux généralistes/polyvalents** existants tels que Apple Siri³¹, Amazon Alexa³², Google Now/Home/Assistant³³, Microsoft Cortana³⁴, donnent des indications, contrôlent les appareils, trouvent des restaurants ou passent des appels. Les assistants virtuels peuvent répondre à des questions sur les sites Web des entreprises et s'interfacer avec des robots. De plus, la technologie des assistants vocaux est incontournable pour les enceintes connectées, pour lesquelles il est l'unique mode d'interaction.³⁵ Comme souligné dans l'Enquête sectorielle sur l'internet des objets pour les consommateurs de l'Union Européenne³⁶, les assistants vocaux permettent la communication entre les composantes matérielles (ex. les enceintes connectées ou les smartphones) et logicielles (les systèmes d'exploitation tels qu'Android de Google et iOS d'Apple pour les dispositifs mobiles intelligents) et augmentent leur complémentarité. En plus, « ces plateformes technologiques permettent également un accès centralisé aux produits et services qu'elles intègrent et le contrôle de ces derniers ».

D'autre part, les **chatbots** sont des systèmes conçus pour des conversations prolongées, qui imitent les conversations non structurées ou les « chats » caractéristiques de l'interaction humain-humain, principalement à des fins de divertissement, mais aussi à des fins pratiques, comme rendre plus naturels les assistants virtuels. C'est le cas par exemple de Blender Bot de Meta-Facebook³⁷, un chatbot fondé sur des méthodes d'apprentissage machine, capable de mener des conversations très naturelles, par exemple, sur du contenu musical, ou le système

³⁰ D. Jurafsky, J. skill. Martin, « Speech and Language Processing », *Prentice Hall*, 2021.

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*, 2019.

³⁶ Commission Européenne, *Rapport final - Enquête sectorielle sur l'internet des objets pour les consommateurs*, 2022, point 164 et s.

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/>)

Xiaolce de Microsoft³⁸, un système qui discute avec des personnes sur des plateformes de messagerie texte. Certains chatbots ont une personnalité plus développée que d'autres, et peuvent également offrir du divertissement et pas seulement une assistance pour les tâches quotidiennes ; ces chatbots sont appelés **chatbots sociaux**.

Histoire des agents conversationnels

Les agents conversationnels ont évolué avec l'augmentation progressive des capacités informatiques et les progrès des outils et techniques de Traitement Automatique du Langage Naturel (TALN). La première mise en œuvre d'un agent conversationnel, qui s'appuyait fortement sur des règles linguistiques encodées à la main, a été réalisée en 1966 avec le développement d'ELIZA³⁹. ELIZA a été conçue pour simuler un psychologue Rogerien, une branche de la psychologie clinique dont les méthodes consistent à faire parler le patient en lui renvoyant ses déclarations. Si un patient dit "J'ai fait une longue promenade en bateau" et que le psychiatre lui répond "Parlez-moi des bateaux", on ne suppose pas qu'il ne sait pas ce qu'est un bateau, mais plutôt qu'il a un objectif de conversation. Cet agent conversationnel peut communiquer avec l'utilisateur par le biais d'un algorithme qui trouve une correspondance entre les mots-clés utilisés dans la demande de l'utilisateur et celles présentes dans les règles utilisées pour reformuler l'entrée et fournir une sortie, c'est-à-dire une réponse à l'utilisateur. Néanmoins, le champ de connaissances d'ELIZA était limité car il dépendait d'une identification minimale du contexte et, en général, les règles définies ne sont pas flexibles pour être facilement mises en œuvre dans de nouveaux domaines. Quelques années après ELIZA, un autre agent conversationnel axé sur la psychologie clinique, PARRY⁴⁰, a été utilisé pour étudier la schizophrénie. En plus des règles de type ELIZA, le système PARRY comprenait un modèle de son propre état mental, avec des variables d'affect pour les niveaux de peur et de colère de l'agent ; certains sujets de conversation pouvaient amener PARRY à se mettre en colère ou à se méfier. Si la variable de colère de PARRY est élevée, il choisira parmi un ensemble de réponses "hostiles". Si l'entrée mentionne son sujet de délire, il augmentera la valeur de sa variable de peur et commencera alors à exprimer la séquence de déclarations liées à son délire. Parry a été le premier système connu à passer le test de Turing (en 1972); les psychiatres ne pouvaient pas distinguer les transcriptions de texte des entretiens avec PARRY des transcriptions d'entretiens avec de vrais paranoïaques.

Dans les années 1980, le domaine des agents conversationnels a connu une importante évolution avec l'utilisation de l'Intelligence Artificielle (AI). Notamment, ChatScript a marqué le début d'une nouvelle ère dans l'évolution technologique des agents conversationnels. L'idée principale de cette technologie innovante est de faire correspondre les entrées textuelles des utilisateurs à un sujet, et chaque sujet est associé à une règle spécifique pour générer une sortie. Il a commencé à déplacer l'attention vers l'analyse et la compréhension sémantiques des phrases. La principale limite de l'utilisation des règles dans les agents conversationnels réside dans le fait qu'elles dépendent du domaine, ce qui les rend inflexibles car elles reposent

³⁸ L. Zhou, J. Gao, Di Li, H.-Y Shum, «The design and implementation of Xiaolce, an empathetic social chatbot », *Computational Linguistics*, 46(1), 2020, pp.53-93.

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.

sur des règles écrites manuellement pour des domaines spécifiques. Avec les progrès récents des méthodes d'apprentissage automatique et de TALN, combinés à la disponibilité de la puissance de calcul, de nouveaux algorithmes ont été créés pour mettre en œuvre des agents conversationnels « avancés » sans s'appuyer sur des règles et des techniques de correspondance de mots, ce qui a encouragé l'utilisation commerciale des agents conversationnels. En particulier, l'application des agents conversationnels s'est étendue avec l'émergence des algorithmes d'apprentissage profond. L'une des nouvelles applications est le développement des assistants vocaux généralistes (comme Alexa d'Amazon, Siri d'Apple, Google Assistant de Google, Cortana de Microsoft et Watson d'IBM). Les assistants vocaux sont généralement intégrés aux smartphones, aux smartwatches, aux enceintes et moniteurs domestiques dédiés, et même aux voitures. Par exemple, lorsque l'utilisateur prononce un mot ou une phrase d'éveil, le dispositif s'active et l'assistant vocal commence à écouter. Grâce à la compréhension de la parole, l'assistant peut alors comprendre les commandes et répondre aux demandes de l'utilisateur, généralement en fournissant des éléments d'information (par exemple, "Alexa, quel temps fait-il aujourd'hui à Paris ?", réponse : "À Paris, le temps est ensoleillé et il fait 75° F"), ou en accomplissant des tâches (par exemple, "Ok Google, joue ma playlist du matin sur Spotify"). Néanmoins, la compréhension de la parole s'est avérée une tâche automatique très difficile en raison des variations tonales, régionales, locales et même personnelles de la parole humaine.

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

Les agents conversationnels interagissent avec les utilisateurs par le biais de ce que l'on appelle le langage naturel, c'est-à-dire le langage humain parlé et écrit. Le traitement automatique du langage naturel (TALN) est la capacité d'un programme informatique à comprendre le langage humain et constitue un domaine de recherche en Intelligence Artificielle (IA). Un agent conversationnel constitue donc une application alimentée par l'IA, qui utilise la reconnaissance vocale, le traitement automatique du langage naturel et la synthèse vocale pour informer, divertir ou fournir des services aux utilisateurs de smartphones et d'enceintes connectées. Techniquement, un agent conversationnel est associé à plusieurs briques d'IA qui répondent aux enjeux suivants : interpréter correctement la demande de l'utilisateur, trouver le résultat approprié et transformer ce résultat en une réponse (écrite ou orale selon le type d'agent conversationnel). Lorsque la réponse est orale, comme dans le cas des assistants vocaux, la formulation de cette réponse est également un enjeu essentiel, tant au niveau de la qualité de l'information qui est fournie comme résultat (c'est-à-dire sa pertinence par rapport à la demande de l'utilisateur) que de la qualité de la réponse (c'est-à-dire la qualité de la formulation de la réponse par une voix synthétique).

Les étapes de l'interaction avec l'utilisateur. L'interaction entre l'utilisateur et l'assistant vocal passe à travers plusieurs étapes⁴¹:

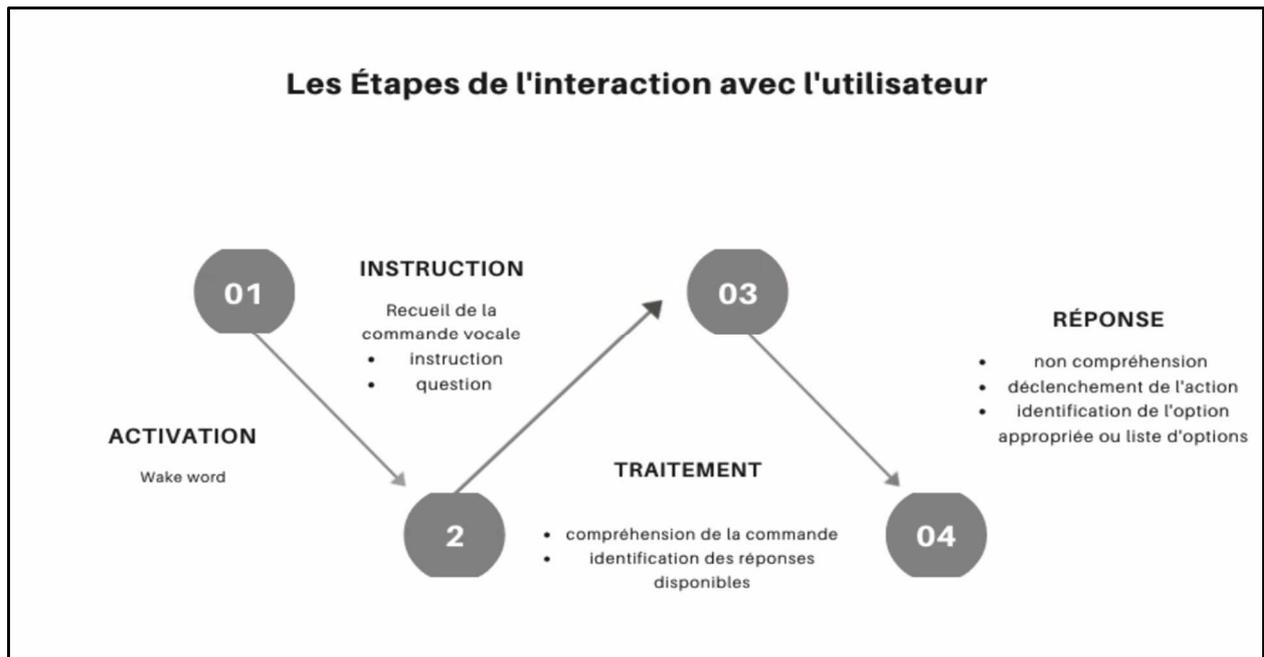
1. L'utilisateur « réveille » l'assistant vocal à l'aide d'un mot clé ou mot d'activation (également appelé « hotword » ou « wake word ») (par exemple, « Dis Siri » pour Apple Siri ou « Ok Google » pour Google Assistant). Cette partie de l'interaction a généralement lieu localement sur le dispositif intelligent. Certains modèles d'assistants vocaux peuvent également reconnaître l'utilisateur grâce à sa voix. Sur certains autres appareils, comme les smartphones, avant de prononcer le mot d'activation, l'utilisateur doit toucher l'écran ou appuyer sur un bouton.
2. L'assistant vocal recueille la commande vocale qui consiste dans une instruction que l'utilisateur lui donne pour exécuter une tâche telle que l'allumage des lumières, la diffusion de musique ou l'envoi d'un message. Il est également possible pour l'utilisateur de poser une vraie question pour recevoir des informations spécifiques telles que les prévisions météorologiques. Cette étape s'appuie sur des algorithmes de reconnaissance de la langue (Natural Language Recognition) qui permettent à l'assistant vocal de reconnaître la langue dans laquelle l'utilisateur s'adresse à lui tout en suivant sa grammaire.
3. L'assistant vocal traite généralement la commande/question reçue par l'utilisateur en deux étapes. La première étape consiste à comprendre la demande et elle est normalement effectuée par une technologie fondée sur le cloud. Cette étape est composée par deux sous-étapes : (1) la reconnaissance vocale (Speech-to-Text) qui permet de traiter la voix de l'utilisateur et de traduire ce signal audio dans une phrase en langage naturel, et (2) la compréhension de la phrase en langage naturel et des intentions exprimées verbalement par l'utilisateur (Natural Language Understanding). L'appareil a la possibilité de conserver un historique des requêtes transcrites, un historique des requêtes audio ou encore d'autres métadonnées associées à la requête telles que l'heure à laquelle cette dernière a été faite. La deuxième étape consiste à identifier les réponses disponibles. L'assistant vocal identifie différentes options pour répondre à la demande de l'utilisateur. Plus précisément, une fois l'intention reconnue l'intention de l'utilisateur avec un indice de confiance suffisant, le programme extrait de la phrase les éléments de contexte nécessaires à l'exécution de la requête, en s'appuyant sur des algorithmes d'extraction d'information, des services tiers (bases de données, recherches Internet, applications, contrôle d'appareils connectés au système) et des éléments de contexte extraits (historique de la conversation et toutes les métadonnées accumulées au fil du temps et des utilisations de l'outil par

⁴¹ Pour une présentation de ces étapes, v. également O. Gouliáeva, E. Dosquet, Y. Moysan, *La révolution des assistants vocaux - Comprendre les enjeux et réussir ses stratégies marketing*. Dunod. 2020, p. 11 ainsi que Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 2022, points 33 à 37.

l'utilisateur). Les réponses disponibles sont classées par l'assistant vocal sur la base de divers paramètres : la pertinence et la disponibilité de la réponse, la satisfaction prévue de l'utilisateur, l'adéquation de la réponse au type de dispositif concerné et l'alignement sur l'activité actuelle de l'utilisateur.

4. L'assistant vocal répond en passant par deux étapes: (1) la synthèse sous forme de phrases de la réponse apportée par la machine à la demande de l'utilisateur avec des algorithmes de génération du langage naturel (Natural Language Generation), et (2) la prononciation de ces phrases automatiquement générées par l'assistant vocal à travers une voix de synthèse, avec des algorithmes de « Text-to-Speech » qui permettent à la machine de transformer un texte en sons en produisant son équivalent parlé et compréhensible par un être humain. Plus particulièrement, il répond selon l'une des options suivantes : (i) faisant savoir à l'utilisateur qu'il ne comprend pas la demande ; (ii) déclenchant une action (ex. jouer une chanson) ou récupérant les informations identifiées comme appropriées (ex. la météo) ; ou (iii) répondant qu'il a identifié l'option la plus appropriée ou fournissant une liste d'options parmi lesquelles l'utilisateur peut choisir celle qui mieux répond à sa demande. Une fois la commande exécutée ou la réponse donnée, si l'utilisateur n'engage pas d'autres actions, l'assistant vocal passe finalement en mode veille.

Il est important de souligner que les algorithmes d'apprentissage automatique employés dans chacune de ces phases d'interaction avec l'utilisateur apprennent de chaque interaction à personnaliser l'expérience de l'utilisateur avec l'assistant vocal. Lorsqu'il est déployé à l'intérieur d'un équipement (téléphone intelligent, enceinte, véhicule), l'assistant vocal est en veille et il est constamment à l'écoute. Toutefois, tant qu'une expression de réveil spécifique n'est pas détectée, aucun message audio n'est transmis à partir de l'appareil récepteur de la voix et aucune autre opération que la détection de l'expression de réveil n'est effectuée.



Interaction humain-machine. L'interaction entre les humains et les systèmes informatiques évolue vers des interfaces fondées sur le langage naturel. Les technologies du traitement automatique du langage naturel permettent aux utilisateurs d'interagir avec le système (ex. système d'exploitation, moteur de recherche, système de question-réponse) dans leurs propres mots, plutôt que d'utiliser l'une des nombreuses façons prédéfinies d'interaction. Dans ce cadre, une conversation textuelle ou vocale est généralement initiée par l'utilisateur qui pose une question en langage naturel et le système répond à la question en langage naturel. Il existe un certain nombre de systèmes qui emploient des interfaces fondées sur le langage naturel à différents niveaux de complexité. Par la suite, une définition de ces systèmes ainsi que de leurs fonctionnalités et leurs différences respectives est fournie.

Interaction avec l'utilisateur. Défis techniques. La compréhension et la génération d'un dialogue avec l'utilisateur est une tâche compliquée. Un dialogue est défini comme une séquence de tours, chacun d'eux étant une contribution unique d'un locuteur au dialogue. Un tour peut consister en une phrase, mais il peut être aussi court qu'un seul mot ou aussi long que plusieurs phrases. La structure des tours a des implications importantes pour le dialogue parlé. Un système doit savoir quand arrêter de parler. Par exemple, si le client s'interrompt, le système doit savoir qu'il faut arrêter de parler (et que l'utilisateur pourrait faire une correction). Un système doit également savoir quand il doit commencer à parler. Par exemple, la plupart du temps, dans une conversation, les locuteurs commencent leur tour presque immédiatement après que l'autre locuteur a terminé, sans longue pause, car les personnes sont capables (la plupart du temps) de détecter lorsque leur interlocuteur est sur le point de finir de parler. Les assistants vocaux doivent également détecter si un utilisateur a fini de

parler, afin de pouvoir traiter l'énoncé et répondre. Cette tâche, appelée *détection du point final*, peut s'avérer très difficile en raison du bruit et du fait que les personnes font souvent des pauses au milieu de leur tour de parole. Une idée clé de la conversation, due à l'origine au philosophe Wittgenstein,⁴² est que chaque énoncé dans un dialogue est une sorte d'action effectuée par le locuteur. Ces actions sont communément appelées *actes de parole* ou *actes de dialogue* (*speech acts*). Les 4 classes principales⁴³ sont :

- Les constatifs : l'engagement du locuteur à ce que quelque chose soit réalisé (répondre, affirmer, confirmer, nier, désapprouver, déclarer) ;
- Les directives : les tentatives du locuteur pour amener le destinataire à faire quelque chose (conseiller, demander, interdire, inviter, ordonner, demander) ;
- Les commissives : engager le locuteur dans une action future (promettre, planifier, jurer, parier, s'opposer) ;
- Les remerciements : exprimer l'attitude du locuteur à l'égard de son interlocuteur en ce qui concerne une action spécifique (s'excuser, saluer, remercier, accepter une reconnaissance).

Ces 4 actes de parole sont utilisés par les assistants vocaux généralistes et les chatbots. Un utilisateur qui demande à un assistant vocal de faire quelque chose (par exemple « Monte le son ») émet une *directive*. Poser une question qui nécessite une réponse est également une façon d'émettre une directive : lorsque le système dit « *Quel jour de mai voulez-vous voyager ?* », c'est comme si le système commandait à l'utilisateur de répondre. En revanche, un utilisateur qui énonce une contrainte (comme « *je dois voyager en mai* ») émet un *constat*. L'acte de parole exprime une composante importante de l'intention du locuteur de dire ce qu'il a dit.

Parfois, la conversation est entièrement contrôlée par un seul participant. Par exemple, un journaliste qui interviewe un chef cuisinier peut poser des questions, et le chef y répond. Dans ce cas, le journaliste a l'initiative de la conversation⁴⁴. Cependant, dans un dialogue humain-humain normal, il est plus courant que l'initiative passe d'un participant à l'autre, tantôt en répondant aux questions, tantôt en les posant, tantôt en orientant la conversation vers de nouvelles directions, tantôt en prenant l'initiative. Ce type d'interaction est dénommée initiative mixte. L'initiative mixte, bien qu'elle soit la norme pour les conversations entre humains, est très difficile à réaliser pour les assistants vocaux et les chatbots. Il est beaucoup plus facile de concevoir des assistants vocaux et des chatbots pour qu'ils soient des répondeurs passifs. Dans les systèmes de question-réponse, ou dans les moteurs de recherche, l'initiative revient entièrement à l'utilisateur. Dans ces systèmes à l'initiative de l'utilisateur, ce dernier spécifie une requête et le système lui répond. L'utilisateur peut ensuite spécifier une autre requête. Par ailleurs, ces systèmes risquent d'enfermer l'utilisateur à qui le chatbot ou assistant vocal pose une question sans lui donner la

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.

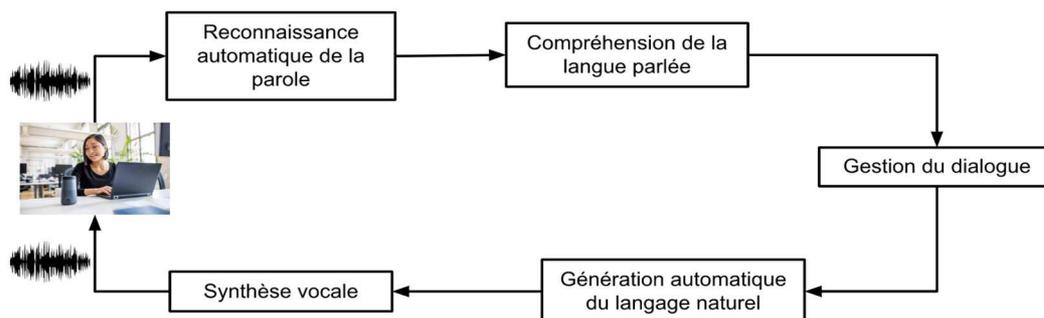
⁴⁴ M. Walker, S. Whittaker, « Mixed initiative in dialogue: An investigation into discourse segmentation », *ACL*, 1990.

possibilité de faire quoi que ce soit avant d'y avoir répondu. Ces systèmes peuvent être très frustrants pour l'utilisateur.

En résumé, l'interaction de l'assistant vocal avec l'utilisateur passe par deux étapes principales, à savoir le traitement et la compréhension de la requête de l'utilisateur et l'identification et la formulation de la réponse. C'est pourquoi, nous nous concentrons sur les enjeux soulevés par ces deux étapes : la requête utilisateur tout d'abord (1) et la génération de la réponse ensuite (2).

1. La requête utilisateur (Speech to Text + TAL)

La requête de l'utilisateur procède de deux phases techniquement bien distinctes se traduisant par un ensemble de procédés vocaux et textuels qui se matérialisent par un système algorithmique de « Speech-to-Text » puis un système de « Text-to-Speech ». Le « Speech-to-Text » permet de convertir de la parole en texte en langage naturel tandis que le « Text-to-Speech », ou synthèse vocale, permet de convertir du texte en une parole synthétisée. Il est possible d'utiliser des voix neuronales préétablies de type humain ou créer une voix neuronale personnalisée. Autrement dit, la première étape de l'interaction de l'assistant vocal avec l'utilisateur suppose l'activation du terminal et l'écoute de la requête (a) et la seconde concerne la compréhension de la requête (b). Ces différentes étapes mobilisent des voix, qu'elles soient humaines ou de synthèse, saisies par des dispositions qui peuvent relever du droit des données personnelles, du droit de la propriété littéraire et artistique, ou plus largement des droits de la personnalité.



a. Activation et écoute

Phase d'activation et écoute. L'assistant vocal est en sommeil jusqu'à son activation par la voix de l'utilisateur, qui va prononcer le mot déclenchant l'activation, ou par une routine qui est soit pré installée, soit programmée par l'utilisateur et dont le déclenchement peut se faire également par la voix. Ce déclenchement par la voix

permet d'améliorer l'interaction avec l'utilisateur et son expérience avec l'assistant vocal, même s'il n'a pas pleinement conscience des effets de l'activation. L'assistant vocal se retrouve au cœur du domicile (véhicule compris) et prend part dès son activation à la vie du foyer par une intrusion dans la sphère de l'intime⁴⁵.

Dès l'activation, les données de contexte (discussions entre les membres du foyer, bruits de fond) font l'objet d'une captation par l'assistant vocal. Cette captation n'est pas neutre et l'on perçoit d'emblée l'enjeu lié à l'information et au consentement des membres du foyer, et celui lié à leur possible enfermement dans cet espace contraint. Le consentement de l'utilisateur et des membres du foyer est largement questionné dans ce cadre, et à l'occasion du déclenchement des routines. Ces dernières permettent d'effectuer une succession de tâches quotidiennes prédéfinies, parfois automatisées, déclenchées directement par la voix de l'utilisateur ou en différé dès lors qu'elles ont été programmées. L'assistant vocal réagira alors à une routine prédéfinie en prononçant le « wake word » (ex : « Bonjour »). À partir de ce moment-là, l'assistant vocal peut, à l'intérieur du domicile, désactiver le mode silencieux, régler les lumières, les prises, les thermostats, donner la météo, les informations sur le trajet domicile-travail, l'agenda, diffuser des contenus (musique, actualités, radio, podcast, livre audio, régler le volume. Ces routines sont déjà configurées sur l'assistant vocal, ce qui n'empêche pas l'utilisateur d'en configurer d'autres, davantage personnalisées⁴⁶. Les différentes interactions alimentent le profil de l'utilisateur en mettant en perspective ses habitudes de vie, ses goûts culturels, ses achats, pouvant contribuer ainsi à son enfermement, outre l'utilisation de ces informations à des fins de ciblage publicitaire ce qui est saisi par les textes européens nouvellement adoptés (v. *infra*), ou à des fins malveillantes si l'installation de l'utilisateur présente des vulnérabilités.

Écoute et données à caractère personnel. La voix de l'utilisateur et celles faisant l'objet d'une captation dans son environnement répondent à la qualification de données à caractère personnel au sens de l'article 4 du RGPD lorsque la personne peut être identifiée soit par le contenu du message qu'elle porte soit en tant que telle dans la mesure où l'on peut déduire de la voix son identité, âge, sexe, origine socioculturelle, état de santé voire son état émotionnel⁴⁷. Le signal vocal émis par la personne est un élément de l'interaction sociale qui catégorise celle-ci selon leur qualification. La voix peut en outre être qualifiée de donnée biométrique au sens de l'article 4 paragraphe 14 du RGPD⁴⁸. Par conséquent, le traitement de la voix de

⁴⁵ « La voix véhicule en dehors du discours (les mots proprement dits) des caractéristiques nombreuses de l'individu : émotions, intentions, condition physique, etc. En se reposant sur des mécanismes de perception, nos interlocuteurs sont en mesure d'interpréter ces signaux et de décrypter ces états (...) Le signal vocal permet donc l'extraction d'informations nombreuses et diverses » : CNIL, À votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux, coll. Livre blanc, n°1, septembre 2020, p. 4.

⁴⁶ HADOPI / CSA, Etude préc.p.17.

⁴⁷ CNIL, Livre blanc préc., p. 5.

⁴⁸ RGPD, art. 14 : on entend par données biométriques « les données à caractère personnel résultant d'un traitement technique spécifiques, résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, qui permettent ou confirment son identification unique, telles que les images faciales ou les données dactylographiques » - en ce sens, CEPD, Lignes directrices sur les assistants vocaux virtuels, 02/2021, version 2.0, 7 juillet 2021, point 31.

l'utilisateur lors des phases d'activation et d'écoute est soumis aux dispositions du RGPD et de la Loi informatique et libertés dès lors que celles-ci trouvent à s'appliquer. On renverra sur ce point à la description présentée par la CNIL dans son livre blanc « A votre écoute », s'agissant du respect des exigences résultant de ces textes⁴⁹ : 1) définition des données à caractère personnel traitées, 2) définition du traitement et de ses finalités, son responsable et sa base légale, 3) choix des données collectées dans le respect des principes d'exactitude, de proportionnalité et de minimisation et ainsi que de leur durée de conservation, 4) information des personnes, exercice et garantie de leurs droits, 5) choix des mesures appropriées pour garantir la confidentialité, l'intégrité et la disponibilité des données à caractère personnel dans le cadre d'une information concise, claire et transparente, fonction de la sensibilité des données et des risques identifiés. Diverses analyses et recommandations ont également été formulées par le Comité européen à la protection des données dans ses lignes directrices sur les assistants vocaux virtuels afin d'accompagner les parties prenantes dans leur mise en conformité⁵⁰. Dans le contexte du présent rapport, certains points d'attention doivent être précisés. Le premier tient aux difficultés liées à l'information de l'utilisateur ; le second, à l'étendue de l'écoute possible.

Interface vocale et information de l'utilisateur. La principale spécificité de l'assistant vocal consiste en son interface avec l'utilisateur à travers des commandes vocales, dépourvue la plupart du temps d'écran ce qui peut faire obstacle de prime abord à son information au sens de l'article 5 du RGPD en l'absence de tableau de bord, ou d'applications dédiées. Rappelons que l'article 5, paragraphe 1, point a) comme les articles 12 et 13, éclairés par le considérant 58 du RGPD, obligent les responsables du traitement à informer les utilisateurs du traitement de leurs données à caractère personnel sous une forme concise, transparente, intelligible et aisément accessible. C'est alors autour de l'information de l'utilisateur que va se situer le nœud du problème dans la mesure où la configuration et la nature de l'interface de l'assistant vocal ne permettent pas nécessairement un bon niveau d'information, ce qui peut remettre en cause la licéité du traitement des données. D'autant que les fournisseurs d'assistants vocaux, et plus largement les responsables de traitement, se heurtent à plusieurs autres obstacles pour se conformer aux exigences de transparence du RGPD⁵¹ qui tiennent à la complexité des flux de données générés et traités. La diversité de nature des données traitées constitue une première difficulté alors que, comme le souligne le CEPD, « *les informations fournies à l'utilisateur doivent correspondre à la collecte et au traitement exacts qui sont effectués* ». Il remarque à cet égard que, « *si certaines métadonnées sont contenues dans un échantillon vocal (par exemple, niveau de stress du locuteur), il n'est pas automatiquement clair qu'une telle analyse est effectuée. Il est essentiel que les responsables du traitement fassent preuve de transparence quant aux aspects spécifiques des données brutes qu'ils*

⁴⁹ CNIL, Livre blanc préc., p. 47-48.

⁵⁰ CEPD, Lignes directrices sur les assistants vocaux virtuels, 02/2021, préc.

⁵¹ Sur ce point, v. CEPD, Lignes directrices relatives à l'article 25 sur la protection des données dès la conception et par défaut, 4/2019, version 2.0 20 octobre 2020.

traitent »⁵². Mentionnons en outre que les fonctionnalités de l'assistant vocal peuvent être supérieures aux attentes des utilisateurs, et ce n'est que la première génération de ce type d'interface. Les assistants vocaux sont davantage intrusifs dans la captation de catégories de données contextuelles et émotionnelles, lors de l'interaction avec l'utilisateur. A cette première difficulté s'ajoute celle tenant à la multitude d'acteurs de l'écosystème de l'assistant vocal, ce qui suppose de porter une attention particulière aux dispositifs permettant l'ajout de fonctionnalités tierces (applications ou *skills*). Si certaines informations générales peuvent être fournies à l'utilisateur lorsqu'il ajoute cette fonctionnalité selon ses propres choix, les frontières entre les différents responsables du traitement concernés peuvent être plus opaques dans le cadre d'une utilisation normale de l'appareil. Comme le souligne le CEPD, « *il est possible que l'utilisateur ne soit pas suffisamment informé sur les personnes qui traitent ses données, sur la manière dont elles le sont (et sur la mesure dans laquelle ce traitement se produit) pour une requête spécifique* »⁵³.

Le CEPD préconise de délivrer à l'utilisateur une information adaptée de manière à garantir la transparence par un meilleur dialogue entre l'humain et la machine. Quant aux modalités formelles de l'information, les lignes directrices rappellent que, conformément à l'article 12, paragraphe 1, du RGPD, la communication orale des informations nécessaires n'est possible que si la personne concernée en fait la demande, mais ne peut être employée comme méthode par défaut. En outre, lorsqu'ils utilisent un message audio pour informer les personnes concernées, les responsables du traitement doivent fournir les informations nécessaires d'une manière concise et claire avec une possibilité de réécoute. Si l'interface vocale reste pertinente pour informer les utilisateurs de l'utilisation de leurs données, notamment pour des utilisateurs souffrant de handicap, le CEPD précise toutefois que l'information de l'utilisateur devra également se matérialiser au moyen d'une application ou d'un service de messagerie⁵⁴. Il ajoute que le moment choisi pour informer l'utilisateur est également déterminant (avant le traitement de données et via une interface personnelle, comme le smartphone). Dans l'hypothèse où le fournisseur d'assistant vocal inclut des applications tierces dans la configuration par défaut - ce qui est majoritairement le cas des assistants vocaux généralistes -, le CEPD relève que celui-ci doit veiller à ce que les utilisateurs obtiennent également les informations nécessaires quant au traitement des données par des tiers. *A fortiori* lorsque l'assistant vocal polyvalent ou spécialisé gère les données d'un ensemble de parties prenantes car il embarque des services et produits (par exemple, télécommunications, commerce électronique, technologies de l'information ou activités web), l'utilisateur doit disposer d'une information claire pour lui permettre de choisir si les données seront utilisées ou non pour établir un profil. Au-delà, il est possible de considérer que la reconnaissance d'un droit au paramétrage - dont la consécration sera envisagée

⁵² CEPD, Lignes directrices sur les assistants vocaux virtuels préc., point 62.

⁵³ CEPD, Lignes directrices sur les assistants vocaux virtuels préc., point 64.

⁵⁴ CEPD, Lignes directrices préc., point 57.

dans la III^{ème} partie de la présente étude - pourrait constituer une solution au traitement de données par ricochet.

Quid de l'écoute en continu ? Il peut apparaître légitime pour le développeur d'un assistant vocal de vouloir accéder à des informations relatives aux interactions que des utilisateurs ont avec son assistant vocal, notamment à des fins d'évaluation et d'amélioration continue du produit. C'est en effet uniquement en conditions réelles (ou de production) que certains comportements du système pourront être observés, un développeur ne pouvant exposer son assistant vocal à l'intégralité des situations susceptibles d'advenir lors de la phase de développement. L'accès à de telles données doit toutefois être rigoureusement encadré, y compris lorsqu'il s'agit de réaliser des écoutes afin d'améliorer le système d'interaction humain-machine. Pour ce faire, et tout particulièrement, l'activation de l'assistant vocal doit être réalisée de façon claire et univoque. Le CEPD souligne ainsi la nécessité pour les utilisateurs de pouvoir déterminer à tout moment si l'assistant vocal « *est actuellement à l'écoute, en circuit fermé et, en particulier, s'il envoie des informations en flux continu en arrière-plan* »⁵⁵ et la nécessité de passer par une confirmation de la réception de la commande vocale (lumière visible, icônes, tableau de bord, etc.)⁵⁶. A cet égard, on se souvient que l'été 2019 a été émaillé de révélations concernant les pratiques de grands acteurs du numérique ayant recours à l'écoute de ces interactions par des employés et prestataires et aux modalités d'accès aux enregistrements. Le Commissaire à la protection des données de Hambourg avait adressé à la société Google une injonction de cesser toutes ses activités d'écoute d'interactions d'utilisateurs avec un assistant vocal par des opérateurs humains⁵⁷. Par la suite, toutes les grandes entreprises ayant mis en œuvre des activités similaires (Apple, Amazon, Microsoft et Facebook notamment) ont mis en suspens ces activités, le temps pour elles de mettre en place des mécanismes de recueil du consentement des utilisateurs satisfaisant au regard de la réglementation⁵⁸. Ces modalités d'information et d'expression du consentement devront être suffisamment explicites et respecter le principe de protection des données par défaut consacré par l'article 25, paragraphe 2 du RGPD⁵⁹.

L'écoute étendue. Si certains sont déployés sur des équipements exclusivement personnels, de nombreux assistants vocaux sont amenés à évoluer dans des environnements partagés. A ce titre, en fonction des lieux dans lesquels évoluent physiquement les assistants vocaux, peuvent se poser des questions relatives à l'exploitation des données de personnes non informées, voire au risque de violations

⁵⁵ CEPD, Lignes directrices préc., point 63.

⁵⁶ CEPD, Lignes directrices préc., point 51.

⁵⁷ https://datenschutz-hamburg.de/assets/pdf/2019-08-01_press-release-Google_Assistant.pdf

⁵⁸ V. CNIL, Livre blanc préc., p. 41.

⁵⁹ A titre d'illustration, on citera ici, dans une autre hypothèse, la condamnation de la société Discorde par la CNIL pour avoir paramétré son application afin qu'elle reste active « *même lorsque l'utilisateur ferme la fenêtre principale, ce qui permet de continuer à communiquer vocalement tout en n'occupant plus de place sur le bureau de l'ordinateur* » et alors que « *Seul un petit indicateur permet de comprendre que l'application est active* ». La CNIL retient alors que l'absence d'information suffisamment claire et visible présente des risques importants pour les utilisateurs, notamment d'intrusion dans leur vie privée, et constitue dès lors une violation de l'article 25 §2 du RGPD : CNIL, Délibération SAN 2022-020, 10 novembre 2022, Discorde INC.

de secrets personnels ou professionnels. Ici encore, les modalités relatives à l'activité d'écoute doivent donc être précisément encadrées. Le CEPD souligne ainsi la nécessité pour le responsable de traitement de « rendre transparent le type d'informations que le VVA peut obtenir au sujet de son environnement » (par ex. l'écoute d'autres personnes se trouvant dans la pièce ou de la musique de fond jouée) ainsi que de respecter son obligation d'information à l'égard de tous les utilisateurs, et non pas seulement l'utilisateur enregistré via son compte dédié, même ceux écoutés à titre accidentel comme les membres du foyer présents au domicile lors de l'activation. Ces exigences seront renforcées dans l'hypothèse où l'assistant vocal fonctionne selon un processus de reconnaissance vocale. La reconnaissance vocale vise en effet à identifier l'utilisateur de manière unique sur la base de sa voix pour activer l'assistant vocal, et suppose dans le même temps de soumettre d'autres personnes (par ex. du foyer) au traitement de leurs données biométriques pour comparer leur modèle vocal par rapport à celui de l'utilisateur préenregistrée qui souhaite être reconnu. En somme, cela implique, lors de la phase d'activation de l'assistant vocal, de filtrer les voix (par ex. des personnes du foyer) pour en identifier une. Dans la mesure où il s'agit alors d'un traitement de données biométriques soumis aux exigences de l'article 9 du RGPD, seront alors requis le consentement explicite de toute personne dont les données doivent être traitées, la réalisation d'une étude d'impact et une transparence accrue s'agissant des effets de la collecte.

Et demain les chatbots : données sensibles, données d'émotions et risques de manipulation. La transcription automatique de la parole⁶⁰ peut révéler des données sensibles que l'on ne peut en principe traiter sauf si le consentement de l'utilisateur, plus exactement de la personne concernée, est acquis conformément aux exigences de l'article 9 du RGPD ; cela exclut que celui-ci puisse être déduit de la simple interaction avec l'utilisateur⁶¹. En revanche, toute la question est de savoir si la voix est en elle-même une donnée sensible ce qui représente un enjeu considérable pour appréhender la nouvelle génération de chatbot dont les fonctionnalités sont étendues. Rien ne s'y oppose, si l'on envisage la question sous l'angle des données biométriques permettant d'identifier la personne de façon unique dès lors que ses caractéristiques physiques, physiologiques ou comportementales la caractérisent de manière définitive. Les données biométriques sont des données sensibles également car elles peuvent être captées à distance à l'insu de la personne. Pour la CNIL, cet élément d'identification du « locuteur » doit être encadré⁶². Des expérimentations ont été autorisées sur des dispositifs de reconnaissance de la voix par des établissements bancaires dans le cadre de serveurs interactifs⁶³.

⁶⁰ A noter que les extraits de données vocales ne sont pas partagés entre les différents acteurs, mais le contenu peut en soi caractériser des données sensibles. Dès lors, l'utilisation se réalise lors de la transcription textuelle des mots, phrases prononcées et mouvements captés, de manière à interpréter le sens du message transmis.

⁶¹ Sur ce point, v. 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, Biométrie, <https://www.cnil.fr/fr/biometrie>

⁶³ CNIL, *A votre écoute*, Livre blanc, préc., p.9.

Au-delà, la voix peut permettre d'inférer des informations relevant de l'état émotionnel de la personne, et certaines technologies sont désormais brevetées afin d'adapter les services proposés en conséquence⁶⁴. C'est le cas notamment de Spotify, avec son propre assistant vocal. Ce service de streaming a obtenu un brevet⁶⁵ sur de nouvelles façons de rendre un assistant vocal plus réactif aux émotions humaines. Le brevet décrit un assistant vocal pour un « appareil de lecture multimédia » qui pourrait, par exemple, reconnaître quand un utilisateur a l'air triste et accuser réception avec sympathie (« vous semblez un peu déprimé ») ou offrir des encouragements (« remonter le moral »). Si un utilisateur semble pressé, l'assistant vocal pourrait essayer de suivre le rythme en parlant plus vite ou en répondant avec moins de mots. D'autres assistants vocaux, comme Alexa, s'orientent également vers l'empathie⁶⁶. La reconnaissance des émotions a un large éventail d'applications pour cet assistant vocal : elle peut rendre les systèmes d'IA conversationnels plus attrayants et fournir un retour d'information implicite des clients qui pourrait aider les agents vocaux comme Alexa à apprendre de leurs erreurs. En outre, les émotions simulées par Alexa offrent aux créateurs d'applications vocales la possibilité de faire en sorte que l'assistant vocal paraisse heureux et excité ou triste et empathique.

Cela conduit alors à s'interroger sur la nature de telles données - que l'on qualifiera de « données émotionnelles » ou de « données d'émotion » - et l'encadrement de leur traitement dès lors qu'ils pourraient s'avérer particulièrement intrusifs. Les données d'émotion permettent de singulariser la personne et composent un corpus de données propres à son identité physiologique et psychique ce qui permet *a minima* de les qualifier de données à caractère personnel au sens de l'article 4 du RGPD⁶⁷. Quant à considérer qu'il s'agit de données sensibles, la lecture de l'article 9 permet seulement de l'envisager dans certains cas de figures, dès lors notamment que la donnée d'émotion relève de la qualification de donnée de santé, se rapporte à la vie sexuelle, ou relève de la catégorie des données biométriques. En dehors de ces hypothèses, les signaux d'émotions ne pourraient qu'être soumis au droit commun des données à caractère personnel. On pourrait dès lors admettre que les données d'émotion puissent être traitées afin qu'un assistant vocal joue un contenu adapté aux émotions ressenties par son utilisateur à la condition qu'un tel traitement respecte bien les principes du RGPD, et notamment le principe de proportionnalité. A l'aune de ces principes, on pourrait penser que tout dépend du contenu proposé au regard de la requête adressée. Autrement dit, admettre que l'utilisateur formule une requête ouverte, du type « joue un morceau de jazz », amènerait sans soulever de réelles difficultés l'assistant vocal à pouvoir choisir le morceau en fonction du contexte d'écoute dans lequel se trouve l'utilisateur.

⁶⁴ CEPD, Lignes directrices préc., point 48.

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_D4Db8t1AljtRTW6jSkXcb1ZTFqBvfWVWOEtCSgaQBd9wgcNUyKS5WiDw

⁶⁶ Adde, V. Rozgnic, « Using Adversarial training to recognize speaker's emotions », 21 mai 2019: <https://www.amazon.science/blog/using-adversarial-training-to-recognize-speakers-emotions>

⁶⁷ G29, *Avis sur le concept de données à caractère personnel*, 4/2007, 20 juin 2007, WP136, pp. 7-8.

Néanmoins, il est possible de considérer qu'admettre de telles applications pourrait laisser le champ ouvert à l'autorisation de bien d'autres traitements de données d'émotion ayant un impact plus intrusif au regard de l'autonomie de la personne voire de son intégrité psychique. L'exploitation des émotions humaines pourrait alors révéler des éléments les plus intimes et mettre à nu la vulnérabilité de chacun. Autrement dit, une approche par les usages ne paraît pas adaptée compte tenu de la nature même des données émotionnelles. En atteste tout particulièrement la perspective du déploiement de divers univers virtuels - ou autres métavers - visant à capter plus finement encore nos émotions par des dispositifs comme les casques de réalité virtuelle et à adapter en conséquence et en temps réel l'environnement virtuel au sein duquel l'avatar d'humain évoluerait⁶⁸.

Par conséquent, la protection *a minima* des données d'émotions par le droit des données personnelles peut paraître insuffisante au vu des enjeux sous-jacents, et notamment des risques de manipulation qui pourraient résulter de ces traitements y compris lorsque visent à en évaluer le profil ou à catégoriser les individus. Cela explique pourquoi certains⁶⁹ considèrent nécessaire d'appréhender ces données d'émotion sous l'angle des « *traitements sensibles* », en faisant une lecture appuyée de l'article 5 paragraphe 1 du RGPD aux termes duquel les données « doivent être : a) traitées de manière licite, loyale et transparente au regard de la personne concernée (licéité, loyauté, transparence) ; b) collectées pour des finalités déterminées, explicites et légitimes », tout en étant « adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées ». Cela pourrait imposer « *une limitation des traitements en cas de risques disproportionnés – contrôle de proportionnalité – mais également une minimisation des données traitées ainsi qu'une véritable légitimité dans les finalités du traitement* »⁷⁰, de manière à éviter les traitements trop invasifs. Mais encore, il faudrait soumettre ces traitements de données à l'article 35 du RGPD qui prescrit une analyse d'impact pour les traitements présentant « *un risque élevé pour les droits et libertés des personnes physiques* », et plus précisément tout « *traitement, en particulier par le recours à de nouvelles technologies, et compte tenu de la nature, de la portée, du contexte et des finalités du traitement, [qui] est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes physiques* » et en particulier, ceux consistant en une « *évaluation systématique et approfondie d'aspects personnels concernant des personnes physiques [...] sur la base de laquelle sont prises des décisions produisant des effets juridiques à l'égard d'une personne physique ou l'affectant de manière significative de façon similaire* »⁷¹. Pour soumettre ces traitements sensibles à analyse d'impact, encore faut-il, selon les critères posés par

⁶⁸ Sur ce point, v. en particulier A. Basdevant, C. François et R. Ronfard, *Mission exploratoire sur les métavers*, octobre 2022, p. 85 et p.90&s - Adde, R. Chatellier, « Métavers : réalités virtuelles ou collectes augmentées ? », *LINC* novembre 2021.

⁶⁹ En ce sens, J. Rochfeld et 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, spéc. p. 759&s.

⁷⁰ J. Rochfeld et C. Zolynski, art. préc. spéc. p. 763.

⁷¹ Idem.

le G29 en son temps, que le système algorithmique procède à une évaluation ou à une notation à partir de profilage ou de prédiction, et que l'on considère que les traitements de données d'émotion « *augmentent le risque possible pour les droits et libertés des personnes* »⁷². Cependant quels sont les droits et libertés que l'on peut invoquer à l'appui de ces traitements ? Si l'on voit bien que ce raisonnement suppose une interprétation téléologique du texte, elle n'en demeure pas moins rattachée à la protection des droits et libertés des personnes. L'article 3 de la Charte des droits fondamentaux selon lequel « *chacun a droit à son intégrité physique et mentale* » pourrait être mobilisé. Néanmoins, le risque de manipulation des émotions dans les interactions avec les assistants vocaux pourrait justifier la consécration de nouveaux droits fondamentaux, couplée utilement aux nouvelles interdictions posées par les textes européens reposant sur l'obligation de loyauté, telles que l'interdiction de la publicité ciblée, des interfaces de choix trompeuses, et de transparence à l'égard de l'interlocuteur machine. Il conviendra d'appréhender, à cet égard, les nouvelles finalités de l'interaction humain-machine qui procèdent de la détection des émotions, de la construction du raisonnement sur le fondement de ces informations affectives, mais qui, à terme, pourront générer de nouvelles émotions propres à une interaction poussée avec la machine⁷³. Les agents conversationnels peuvent être dotés de modules de détection des émotions à partir d'indices linguistiques, corporels, et de modulation dans la voix⁷⁴ contribuant ainsi à développer des stratégies de dialogues adéquates pour être en phase avec les émotions de l'utilisateur⁷⁵. Un enfermement qui peut aller jusqu'au clonage de la voix, et du changement d'émotion de l'utilisateur lorsqu'il interagit avec la machine.

De telles interrogations apparaissent déjà avec le déploiement des premiers Deadbots, ces agents conversationnels susceptibles de figurer une conversation avec une personne morte à partir des données vocales ou textuelles laissées par la personne défunte à l'image de Replika⁷⁶. Le RGPD ne régit pas les données des personnes décédées. Il faut se reporter aux articles 84 à 86 de la loi n°78-17 modifiée du 6 janvier 1978 informatique et libertés pour y trouver des principes d'encadrement. Les droits détenus au titre de cette législation par les personnes vivantes peuvent être maintenus provisoirement pour organiser sa succession et notamment dans deux hypothèses. La première, que l'on identifie comme étant une « mort numérique », permet à la personne de rédiger de son vivant des directives générales ou spéciales à l'endroit de personnes différentes. Cette hypothèse reste le plus souvent théorique. Reste la seconde, faisant état de l'absence d'anticipation de la personne qui n'aura pas scellé le sort de ses données. Cette prolongation virtuelle n'est pas sans

⁷² G29, *Lignes directrices concernant l'analyse d'impact relative à la protection des données (AIPD) et la manière de déterminer si le traitement est « susceptible d'engendrer un risque élevé » aux fins du règlement (UE) 2016/679*, révisées le 4 oct. 2017, WP 248 rév. 01, p. 11, critères 1 et 4.

⁷³ CNPEN, *Agents conversationnels : enjeux d'éthique*, avis n°3, novembre 2021, p. 20.

⁷⁴ L. Devillers, « Le dialogue humain-machine. Intelligence artificielle/intelligence humaine : manipulation et évaluation », *Futuribles* 2019, n° 433, p. 51, spéc. p. 52.

⁷⁵ R. Chatellier, « Captation des émotions : comment vous le direz pourra être retenu contre vous... », *LINC* avril 2018.

⁷⁶ Pour une description de Replika, v. annexe 2 - également, H. Guillaud, « A qui les interfaces s'adaptent-elles ? », *InternetActu*, 20 janvier 2019.

conséquences éthiques et juridiques et en appelle à l'élaboration d'un cadre juridique que l'on devine utile pour les nouveaux responsables de traitement que sont les ayant droits. Plus de dix millions de personnes interagissent ainsi avec Replika dans le monde, et ce n'est que le premier du genre. Disposant des données nécessaires pour régler la succession du défunt, l'ayant droit a accès aux données personnelles et à toutes les informations utiles à la liquidation et au partage de succession, telles que les photos. Ce qui permet d'alimenter un Deadbot mais également un avatar. Il n'en demeure pas moins que ce dispositif est transitoire et répond à l'exécution des directives du défunt. Dès lors, il ne peut justifier la création d'un Deadbot ou d'un avatar qui s'installerait dans le temps à la place du défunt. Un droit à l'oubli spécifique, intégré « by design », que l'on pourrait concevoir à titre de deuil est probablement à circonscrire de manière à éviter une prolongation virtuelle dès le décès. L'enjeu est de taille. Judith Rochfeld précise à cet effet que « *l'impératif d'oubli pourrait ainsi être croisé à l'organisation sociale du temps de la mort. Les solutions pourraient alors ne plus dépendre des seules forces du marché et des volontés des individus (ou de leurs proches) mais d'un statut social du deuil face à des forces numériques « éternelles » : d'une considération juridique de la dimension techno-sociale de la mort* »⁷⁷. Par ailleurs, le recours à la protection de la dignité humaine ne doit pas être écarté dès lors qu'il s'agira de protéger des « bribes de personnalité » de la personne décédée⁷⁸, sous réserve de rapporter la preuve de leur existence.

b. Compréhension de la requête utilisateur

Compréhension de la requête de l'utilisateur et données d'entraînement. La requête de l'utilisateur est interprétée à l'aide des technologies de traitement du langage naturel. Les intentions du message sont extraites et des variables d'information sont identifiées pour permettre de retrouver le patron linguistique le plus pertinent pour représenter la requête et donc lui associer une réponse pertinente. L'un des enjeux pour cette étude consiste à envisager comment les algorithmes de TALN conçus pour comprendre la requête de l'utilisateur ont été entraînés et avec quelles données. Les agents conversationnels fondés sur des corpus (c'est-à-dire sur des jeux de données textuelles et de la parole), au lieu d'utiliser des règles construites à la main pour générer les réponses, exploitent des réelles conversations entre humains pour entraîner des assistants vocaux ou des chatbots. Ces systèmes nécessitent des centaines de millions, voire des milliards de mots pour l'entraînement⁷⁹. Les jeux de données disponibles comprennent des transcriptions de conversations en langage naturel, comme le corpus Switchboard de conversations téléphoniques en anglais

⁷⁷ J. Rochfeld, « Les avatars d'éternité - Vers de nouvelles personnes résiduelles compassionnelles ? », *Mélanges en l'honneur de Catherine Labrusse*, IRJS éd., 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.

américain⁸⁰ ou les divers corpus de conversations téléphoniques CALLHOME et CALLFRIEND, disponibles dans de nombreuses langues. De nombreux systèmes s'entraînent également sur des dialogues de films⁸¹, qui ressemblent à de nombreux égards à des conversations naturelles. Des jeux de données ont également été créés spécifiquement pour l'entraînement des systèmes de dialogue en engageant des *crowdworkers* pour tenir des conversations, en leur faisant souvent endosser des personnages ou parler de connaissances qui leur ont été fournies au préalable. Par exemple, le jeu de données EMPATHETIC DIALOGUES comprend 25000 conversations crowdsourcées fondées sur une situation spécifique où un locuteur ressentait une émotion spécifique⁸². Tous ces ensembles de données, bien que volumineux, n'atteignent pas la taille de milliards de mots ; c'est pourquoi de nombreux systèmes effectuent d'abord un pré-entraînement sur de grands ensembles de pseudo-conversations tirées des réseaux sociaux, comme Twitter⁸³ et Reddit⁸⁴. Une autre technique courante consiste à extraire des réponses possibles de sources de connaissances (telles que Wikipédia ou les articles de presse) afin qu'un chatbot puisse raconter des histoires ou mentionner des faits acquis de cette manière.

Enfin, une fois qu'un agent conversationnel a été mis en pratique, les tournures que les humains utilisent pour répondre au chatbot peuvent être utilisées comme données conversationnelles supplémentaires pour l'entraînement ou un réglage plus fin. Il est important de disposer de métriques de confiance pour s'assurer que ces exemples de conversation proviennent de conversations qui se déroulent bien et non de conversations trompeuses ou nuisibles que l'agent conversationnel ne doit pas apprendre, comme par exemple des conversations haineuses. Il est également crucial dans ces cas de supprimer les informations personnelles identifiables⁸⁵.

Les agents conversationnels peuvent également être construits avec des architectures hybrides, en conjuguant des algorithmes fondés sur des règles et des réseaux de neurones. C'est le cas, par exemple, des systèmes qui concourent au défi Alexa Prize, dans lequel des équipes universitaires construisent des chatbots sociaux pour converser avec des volontaires sur la plateforme Amazon Alexa, et sont notés en fonction de la longueur et des évaluations des utilisateurs de leurs conversations.

Enjeux techniques. Les algorithmes des assistants virtuels, utilisés sous diverses formes dans des systèmes comme Apple Siri, Amazon Alexa et Google Assistant, sont conçus autour d'un cadre fixe. Leur objectif est d'identifier un patron linguistique

⁸⁰ J. Godfrey, E. Holliman, J. McDaniel, « SWITCH- BOARD: 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 dialogs ». *2nd Workshop on Cognitive Modeling and Computational Linguistics, 2011*; et 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 modeling 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.

⁸⁵ CNIL, *À votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*, coll. Livre blanc, n°1, septembre 2020, p. 54.

dans la requête formulée par l'utilisateur pour savoir à quel cadre il faudra faire référence pour chercher une réponse pertinente. Plus précisément, l'objectif du système est de remplir les cases du cadre (ou patron) avec les éléments de remplissage souhaités par l'utilisateur, puis d'exécuter l'action correspondante pour l'utilisateur (c'est-à-dire répondre à une question, écouter une chanson ou réserver un vol). Pour ce faire, le système pose des questions à l'utilisateur (à l'aide de modèles de questions pré-spécifiés associés à chaque slot de chaque cadre), remplissant tout slot spécifié par l'utilisateur. Des exemples de cadre sont les suivantes :

Slot	Type	Cadre de la question
VILLE D'ORIGINE	ville	« De quelle ville partez-vous ? »
DESTINATION	ville	« Où allez-vous ? »
JOUR D'ARRIVÉE	date	« Quel jour voulez-vous arriver ? »

Si la réponse d'un utilisateur remplit plusieurs slots, comme dans le cas suivant : « *Je veux un vol de Paris à Lyon, aller simple, partant après 14h le mardi* », le système remplit tous les créneaux pertinents, puis continue à poser des questions pour remplir les créneaux restants, en sautant les questions associées aux créneaux remplis. Par exemple, une règle associée à l'emplacement DESTINATION du cadre de réservation d'avion, une fois que l'utilisateur a spécifié la destination, peut automatiquement entrer cette ville comme lieu de séjour par défaut pour le cadre de réservation d'hôtel correspondant. Ou, si l'utilisateur spécifie le JOUR DE DESTINATION pour un voyage de courte durée, le système pourrait automatiquement saisir le JOUR D'ARRIVÉE. De nombreux domaines nécessitent plusieurs cadres. Différents types d'entrées provoquent le déclenchement de différentes productions, chacune d'entre elles pouvant remplir de manière flexible différents cadres. Les règles de production peuvent alors changer de contrôle en fonction de facteurs tels que l'entrée de l'utilisateur et un simple historique de dialogue comme la dernière question posée par le système. Lorsque le système dispose de suffisamment d'informations, il exécute l'action nécessaire (comme l'interrogation d'une base de données de vols) et renvoie le résultat à l'utilisateur.

L'objectif de ces algorithmes est d'extraire trois éléments de la requête de l'utilisateur. La première tâche est la classification par domaine : cet utilisateur parle-t-il par exemple de compagnies aériennes, programme-t-il un réveil ou veut-il écouter une chanson ? Cette tâche de classification est inutile pour les systèmes à domaine unique qui se concentrent, par exemple, uniquement sur l'écoute de la musique, mais les chatbots ou les assistants vocaux généralistes sont très impactés par cette tâche. La seconde est l'intention de l'utilisateur : quelle tâche ou quel objectif général l'utilisateur

essaie-t-il d'accomplir ? Par exemple, la tâche pourrait être de trouver un film, d'afficher un vol ou d'écouter une chanson qui lui fasse remonter le moral. Enfin, il est nécessaire de remplir les créneaux : extraire les cadres particuliers que l'utilisateur souhaite que le système comprenne à partir de son énoncé par rapport à son intention.

Text and Data Mining. Les techniques de Text and Data mining (TDM) sont largement utilisées par des entités privées et publiques pour analyser de grandes quantités de données (y compris des contenus protégés au titre du droit d'auteur comme des textes, des images, des vidéos, etc.) dans différents domaines, notamment pour le développement de nouvelles applications ou technologies. L'utilisation de TDM techniques sur des œuvres protégées par le droit d'auteur peut être soumise à différentes conditions, selon le cadre juridique. En termes génériques, un acte de reproduction est requis avant que le TDM puisse être appliqué sur le contenu accessible sur le Web. Quant au bénéfice de l'exception consacré par la directive 2019/790 Droit d'auteur dans le marché unique numérique⁸⁶ transposée dans le Code de la propriété intellectuelle par l'ordonnance du 21 novembre 2021⁸⁷, il est subordonné à l'*opt out* du titulaire de droit dès lors que l'on se trouve hors du cadre de la recherche. L'article 4.3 de la directive, repris par l'article L. 122-5-3-I du CPI, énonce que cet *opt out* doit être expresse et doit s'exprimer « *de manière appropriée, notamment par des procédés lisibles par la machine pour des contenus mis à la disposition du public en ligne* »⁸⁸. L'article R. 122-28 du CPI, introduit par le Décret du 23 juin 2022⁸⁹ vient en préciser les termes et dispose que « *L'opposition mentionnée au III de l'article L. 122-5-3 n'a pas à être motivée et peut être exprimée par tout moyen. Dans le cas de contenus mis à la disposition du public en ligne, cette opposition peut notamment être exprimée au moyen de procédés lisibles par machine, y compris des métadonnées, et par le recours à des conditions générales d'utilisation d'un site internet ou d'un service* ».

Dans ce contexte, une solution lisible par machine qui rationalise la communication des droits et licences TDM disponibles pour les contenus protégés par le droit d'auteur en ligne est nécessaire pour faciliter le développement de ces applications et réduire les risques d'incertitude juridique pour les acteurs qui mettent en œuvre ces techniques. Cette solution vise aussi à optimiser la capacité des acteurs de TDM à accéder licitement à des contenus utiles et à les traiter à grande échelle. Le W3C (World Wide Web Consortium) Community Group « TDM Reservation Protocol »⁹⁰ a récemment proposé un nouveau protocole appelé TDMRep. L'objectif de ce protocole

⁸⁶ Directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

⁸⁷ Ordonnance n°2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

⁸⁸ V. A. Bensamoun et Y. Bouquerel, *Transposition des exceptions de fouille de textes et de données. Enjeux et propositions*, Rapport de mission présenté au CSPLA, 15 décembre 2020.

⁸⁹ Décret n°2022-928 du 23 juin 2022 portant modification du code de la propriété intellectuelle et complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE

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

est notamment de permettre à un titulaire de droits de déclarer son choix concernant la fouille de textes et de données (TDM) sur les ressources Web qu'il contrôle, ce qui permet aux destinataires de cette déclaration d'adapter leur comportement en matière de *scraping* ou de conclure un accord distinct avec le titulaire de droits qui satisfasse toutes les parties. Plus précisément, ce protocole fournit trois techniques complémentaires pour exprimer les choix des titulaires de droits. Celles-ci correspondent à des situations et des compétences techniques différentes que peut avoir un éditeur : (1) le fichier TDM sur le serveur d'origine est un mécanisme permettant de déclarer les choix des titulaires de droits pour l'ensemble du site dans un fichier hébergé sur le serveur d'origine du contenu Web qu'un agent TDM souhaite exploiter; (2) le champ d'en-tête TDM est un mécanisme permettant de déclarer un choix dans une réponse HTTP; (3) les méta-données TDM dans le contenu HTML sont un mécanisme permettant de déclarer un choix intégré dans le contenu HTML.

2. La génération de la réponse (TAL + Text to Speech)

La génération de la réponse par l'assistant vocal passe par trois étapes principales. Tout d'abord, le système doit trouver la réponse la plus pertinente par rapport à la requête de l'utilisateur (a). Ensuite, le système doit générer la réponse en langage naturel, en utilisant des algorithmes de génération en langage naturel (b). Enfin, le système doit exprimer la phrase générée par une voix synthétique (c). Il convient de mettre en évidence les principaux enjeux techniques et juridiques liés à ces trois étapes.

a. La génération de la réponse (TAL)

Le module de génération du langage naturel de l'assistant vocal produit les énoncés que le système répercute à l'utilisateur. La plupart des systèmes sont fondés sur des cadres, c'est-à-dire qu'ils ont tendance à utiliser une génération de modèles, dans laquelle tous ou la plupart des mots de la phrase à prononcer à l'utilisateur sont pré-spécifiés par le concepteur de l'assistant virtuel. Les phrases créées par ces modèles sont souvent appelées *prompts*. Les modèles peuvent être complètement fixes (ex. « Bonjour, comment puis-je vous aider ? ») ou peuvent inclure certaines variables remplies par le générateur dépendant de la requête de l'utilisateur (ex. « Quelle est la météo à Paris aujourd'hui ? », « La météo à NOM_VILLE est MÉTEO » -- « La météo à Paris est nuageuse »). Un composant de génération plus sophistiqué peut conditionner le contexte exact pour produire des tournures qui semblent beaucoup plus naturelles. Les assistants vocaux actuels mettent en place une combinaison des deux approches.

Réponse unique. La plupart des chatbots et des assistants vocaux fondés sur des corpus produisent leurs réponses au tour de parole d'un utilisateur en contexte, soit par des méthodes de *récupération d'information* (en utilisant la recherche d'information pour extraire une réponse d'un corpus approprié au contexte du dialogue), soit par des méthodes de *génération du langage naturel* (en utilisant un modèle de langage ou un algorithme fondé sur un réseau de neurones pour générer la réponse en fonction du contexte du dialogue). Dans les deux cas, les systèmes génèrent en principe une seule réponse appropriée à l'ensemble de la conversation jusqu'à présent. C'est pourquoi on les nomme souvent des systèmes de question-réponse. Les algorithmes de chatbot ou assistants vocaux fondés sur des corpus s'inspirent donc des algorithmes des systèmes de question-réponse, qui se concentrent également sur des réponses uniques tout en ignorant les objectifs conversationnels à plus long terme. Comme souligné dans le Livre Blanc de la CNIL sur les assistants vocaux³, « *le fait de ne fournir qu'une réponse unique soulève des interrogations quant au choix des sources de résultats* ».

Moteur de recherche et moteur de réponse. Le problème de la réponse unique repose sur les algorithmes mis en place par les moteurs de recherche et les systèmes de question-réponse.

Concernant le fonctionnement des moteurs de recherche, la recherche neuronale initiale, supposait des méthodes beaucoup plus simples, comme la correspondance exacte de chaînes de caractères. La correspondance de chaînes de caractères permettait de filtrer un certain terme sur un site Web ou dans un document. La correspondance exacte des chaînes de caractères renvoie à tous les documents qui contiennent le terme recherché, sans ordre particulier. Pour augmenter la pertinence des résultats de recherche, il était possible d'utiliser la correspondance des mots-clés. Selon les méthodes de correspondance des mots clés telles que tf-idf, les termes d'un document sont pondérés en fonction de leur pertinence dans le corpus. Ces méthodes sont rapides et indépendantes de la langue. En revanche, la correspondance par mots-clés fonctionne selon le principe simplificateur du « *bags of words* ». Cela signifie qu'elle ne préserve pas l'ordre des mots dans un document. Or l'ordre des mots est important pour le sens d'une phrase car il contribue à notre compréhension des relations syntaxiques et sémantiques entre les mots. Au sein des moteurs de recherche modernes, des algorithmes de recherche sémantique (ou "neuronale") sont mis en place. A ce titre, au lieu de comparer une requête (par exemple, une chaîne de caractères ou une image) directement à la base de données, elle passe par un réseau neuronal qui a été pré-entraîné sur des millions de données. Le modèle neuronal a appris à coder une requête comme un vecteur à haute dimension. Cette haute dimensionnalité permet aux modèles neuronaux de mieux saisir le sens ou la valeur sémantique, d'une requête. Les derniers modèles neuronaux de langage sont tous fondés sur l'architecture des transformeurs.

Les transformeurs

Un « modèle de transformeur » est une architecture de réseau de neurones qui permet d'effectuer des tâches complexes telles que la génération automatique de texte ou d'images avec des coûts d'entraînement réduits. Notamment, elle est constituée de couches de transformateur capables de modéliser des dépendances séquentielles entre différents éléments (images, pixels, mots) pour réduire le temps d'entraînement de ces modèles. Par exemple, les modèles de transformeur sont utilisés pour résoudre toutes sortes de tâches en traitement automatique du langage naturel. Le pipeline d'un modèle de transformateur permet de connecter un modèle avec ses étapes de pré-traitement et de post-traitement nécessaires. Par défaut, ce pipeline sélectionne un modèle pré-entraîné particulier qui a été affiné pour une tâche NLP précise. Trois étapes principales sont nécessaires lorsqu'on transmet du texte à un modèle de transformeur : (1) le texte est pré-traité dans un format que le modèle peut comprendre ; (2) les entrées pré-traitées sont transmises au modèle ; (3) les prédictions du modèle sont post-traitées, afin qu'on puisse leur donner un sens.

L'architecture de transformeurs a été introduite en juin 2017. Les recherches initiales étaient axées sur les tâches de traduction automatique. Elle a été suivie par l'introduction de plusieurs modèles influents, notamment :

- *Juin 2018* : GPT, le premier modèle de transformeur pré-entraîné, utilisé pour le réglage fin sur diverses tâches de TALN et qui a obtenu des résultats de pointe.
- *Octobre 2018* : BERT, un autre grand modèle pré-entraîné, celui-ci conçu pour produire de meilleurs résumés de phrases.
- *Février 2019* : GPT-2, une version améliorée (et plus grande) de GPT qui n'a pas été immédiatement rendue publique pour des raisons d'éthique.
- *Octobre 2019* : DistilBERT, BART et T5, des grands modèles préformés utilisant la même architecture que le modèle de transformeur original.
- *Mai 2020* : GPT-3, une version encore plus grande de GPT-2, capable d'obtenir de bonnes performances sur une variété de tâches sans avoir besoin d'un réglage fin.

Tous les modèles de transformeur mentionnés ci-dessus (GPT, BERT, BART, T5, etc.) ont été formés comme des modèles de langage. Cela signifie qu'ils ont été formés sur de grandes quantités de texte brut de manière auto-supervisée. L'apprentissage auto-supervisé est un type d'entraînement dans lequel l'objectif est automatiquement calculé à partir des entrées du modèle. Cela signifie que les humains ne sont pas nécessaires pour étiqueter les données d'entraînement. Ce type de modèle développe une compréhension statistique de la langue sur laquelle il a été formé. Un exemple de tâche consiste à prédire le mot suivant dans une phrase après avoir lu les n mots précédents. Il faut cependant être prudents avec ces modèles qui fonctionnent en observant les relations statistiques entre les mots et les phrases, mais qui ne comprennent pas leur sens.⁹¹

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

Quant aux moteurs de réponse, ils interprètent la question, évaluent les options disponibles et fournissent la meilleure réponse possible. Plus précisément, un

⁹¹ <https://www.nature.com/articles/d41586-021-00530-0>

système de réponse aux questions accepte une question et renvoie ensuite non pas un document, mais une réponse. Le terme « réponse extractive à une question » fait allusion au fait que le système met en évidence un ou plusieurs passages (d'un ou plusieurs documents) comme réponses. Un moteur de réponse fournit normalement une réponse unique, qui est celle jugée la plus pertinente parmi les réponses possibles repérées par l'algorithme.

Évaluation. L'évaluation est cruciale dans la conception des chatbots et des assistants vocaux. Ceux-ci sont généralement évalués différemment car ils ont des objectifs différents : les assistants vocaux généralistes doivent accomplir une tâche comme la réservation d'un vol ou jouer une chanson ; les chatbots ont un autre type d'objectif, comme être agréable à l'égard des utilisateurs. Les chatbots sont évalués par des humains, qui leur attribuent une note. Il peut s'agir de l'humain qui a parlé au chatbot (évaluation du participant) ou d'un tiers qui lit la transcription d'une conversation humain/chatbot (évaluation de l'observateur). Dans l'évaluation des participants, l'évaluateur humain discute avec le modèle pendant quelques minutes et note le chatbot sur 8 dimensions en capturant la qualité conversationnelle : éviter les répétitions, l'intérêt, le sens, la fluidité, l'écoute, la curiosité, l'humanité et l'engagement. Les évaluations par des observateurs utilisent des annotateurs pour examiner le texte d'une conversation complète ; par exemple les évaluateurs peuvent noter la cohérence de chaque tour.

b. L'expression de la réponse : la synthèse vocale (Text to Speech)

L'expression de la réponse comporte deux grands enjeux : tout d'abord, une fois que le système a trouvé la réponse la plus pertinente à fournir à l'utilisateur, il formule cette réponse en langage naturel, grâce à des algorithmes de génération de langage naturel. Ce texte est ensuite traduit en un fichier audio à l'aide d'algorithmes de synthèse vocale (*Text-to-Speech*). Le deuxième défi concerne la génération de la voix synthétique, qui fait appel à différents algorithmes de synthèse vocale. Il convient par conséquent d'envisager ces deux défis.

Text-to-Speech. La synthèse vocale ou synthèse de la parole (Text-to-Speech - TTS) vise à synthétiser une parole intelligible et naturelle à partir d'un texte. Elle constitue depuis longtemps un sujet de recherche en IA, en traitement du langage naturel et en traitement de la parole. Le développement d'un système TTS nécessite des connaissances sur les langues et la production de la parole humaine, et implique de multiples disciplines, notamment la linguistique, l'acoustique, le traitement du signal numérique, et l'apprentissage automatique. Un système TTS se compose d'un socle de trois éléments : un module d'analyse de texte, un modèle acoustique et un vocodeur. Le module d'analyse de texte convertit une séquence de texte en caractéristiques linguistiques, les modèles acoustiques génèrent des caractéristiques

acoustiques à partir des caractéristiques linguistiques, puis les vocodeurs synthétisent la forme d'onde à partir des caractéristiques acoustiques.

Pluralité de techniques. Différentes techniques sont proposées pour la génération de la voix de synthèse (WaveNet⁹², Tacotron⁹³, DeepVoice⁹⁴, Google Duplex⁹⁵). La méthode la plus courante pour produire une voix de synthèse est appelée synthèse concaténative. Cette méthode recherche les phonèmes (c'est-à-dire les unités sonores distinctes dans une langue spécifique) et associe ces « morceaux de parole » enregistrés à partir de voix humaines pour produire une parole synthétisée. En général, cette approche ne produit pas une bonne qualité de voix en raison de la manière dont elle assemble ces unités. Le système puise dans une base de données audio, en segmentant les formes d'onde qui créent suffisamment de variations de la parole pour que l'oreille humaine ne les perçoive pas de manière réaliste. Des avancées importantes dans ce domaine ont été proposées avec les réseaux neuronaux. L'utilisation d'un réseau neuronal permet de prendre un ensemble ordonné de phonèmes, puis de le transformer en un ensemble de spectrogrammes.⁹⁶ Le réseau neuronal les utilise et sélectionne les spectrogrammes appropriés avec les bandes de fréquences qui articulent plus précisément les caractéristiques acoustiques que le cerveau humain utilise pour comprendre et organiser systématiquement la parole. Un vocodeur neuronal traduit ensuite ces spectrogrammes en formes d'ondes vocales, ce qui permet de reproduire la voix de manière naturelle. Ces approches permettent aussi de modéliser les paramètres d'une voix comme l'intonation, le rythme, et le timbre, par des auto-encodeurs et des réseaux adverses génératifs.

Il existe trois principales techniques de synthèse concaténative :

- (1) La synthèse spécifique à un domaine : elle consiste à enchaîner des mots ou des phrases préenregistrés pour former des propos complets. Cette méthode est couramment utilisée pour des cas d'utilisation simples et répétitifs tels que les bulletins météo ou les instructions données avant le décollage. Cependant, cette solution est limitée aux mots et aux phrases pour lesquels elle a été programmée.
- (2) Synthèse de sélection d'unités : elle puise dans une vaste base de données de clips audio de discours préenregistrés d'un acteur préalablement transcrits et décomposés en phonèmes, syllabes, morphèmes, mots, expressions et phrases. Ces unités sont ensuite indexées et rassemblées pour déterminer la meilleure séquence pour la phrase cible.
- (3) Synthèse de diphones : exploite une base de données contenant tous les diphones présents dans une langue spécifique.⁹⁷ Une fois tirée, la prosodie est superposée à ces unités à l'aide de techniques de traitement numérique du

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>

⁹⁶ Un spectrogramme est un rendu visuel du spectre des bandes de fréquences d'un signal.

⁹⁷ Par exemple, l'espagnol compte environ 800 diphones alors que l'allemand en compte 2500.

signal. Le résultat n'est pas aussi bon que la sélection d'unités mais meilleur que la synthèse spécifique à un domaine.

Comme souligné par la CNIL⁹⁸, ce sujet est important par rapport aux *deepfake* qui exploitent les réseaux de neurones pour manipuler ou générer des contenus sonores à fort potentiel de tromperie.

Enjeux techniques en Text-to-Speech

Avec le développement de l'apprentissage profond, le TTS fondé sur les réseaux neuronaux (TTS neuronal) est proposé, adoptant les réseaux neuronaux profonds comme modèle de base pour la synthèse de la parole. Un élément clé de la construction d'un modèle de reconnaissance automatique de la parole est de l'entraîner sur une série de voix différentes, afin qu'il puisse apprendre une variété de profils de fréquences acoustiques et différentes manières d'exprimer les phonèmes, qui sont les unités les plus courtes de la parole. Par exemple, le modèle WaveNet pour générer directement la forme d'onde à partir de caractéristiques linguistiques peut être considéré comme le premier modèle neuronal de TTS. D'autres modèles comme DeepVoice s'appuient sur la synthèse paramétrique statistique, mais l'améliore avec les modèles fondés sur les réseaux neuronaux. Certains modèles *end-to-end* (par exemple, Tacotron, Deep Voice et FastSpeech) sont proposés pour simplifier les modules d'analyse de texte et prendre directement en charge les caractères et les phonèmes en entrée, et de simplifier les caractéristiques acoustiques avec des mélo spectrogrammes. Plus précisément, le système de TTS extrait les incorporations de prosodie d'un fichier audio de référence et l'utilise comme entrée additionnelle au réseau de neurones. Pendant l'apprentissage, un fichier audio de référence est utilisé, et pendant l'inférence, un autre fichier audio de référence contribue à synthétiser la parole avec une prosodie similaire. Les tokens de style peuvent augmenter la capacité et la variation des modèles TTS pour apprendre différents types de styles. Chaque token peut apprendre différentes représentations de la prosodie, comme les différents rythmes d'élocution et les émotions. Le modèle peut utiliser un fichier audio de référence pour assister et extraire les représentations de la prosodie, ou simplement choisir un ou plusieurs tokens de style pour synthétiser la parole. Les modèles TTS adaptatifs ont l'objectif de synthétiser la voix d'un utilisateur. Ces modèles sont à la base des systèmes d'adaptation de la voix, de clonage de la voix, de la voix personnalisée. Dans ce cadre, un modèle de TTS source (généralement formé sur un ensemble de données vocales à plusieurs locuteurs) est généralement adapté avec quelques données d'adaptation pour chaque voix cible.

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

Ces modèles sont à l'origine des voix de synthèses des assistants vocaux et des agents conversationnels. Par exemple, Amazon Alexa a synthétisé 32 fois chaque énoncé d'un ensemble de données, en échantillonnant de manière aléatoire 32 profils

⁹⁸ Livre blanc de la CNIL, *A votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*, septembre 2020, p.7.

vocaux parmi les 500 recueillis auprès de volontaires en laboratoire. Certaines données peuvent provenir de Cleo, une compétence Alexa qui permet aux clients multilingues de contribuer à l'apprentissage de nouveaux modèles linguistiques en répondant aux *voice prompts* par des énoncés à forme ouverte. Ces données ne sont pas rendues publiques. Dans le cas d'Apple Siri, pour iOS 11, c'est un talent vocal féminin qui a été choisi dans le but d'améliorer la personnalité et l'expressivité de la voix de Siri. Des centaines de candidats ont été évalués avant d'en choisir une. Ensuite, plus de 20 heures de discours ont été enregistrées pour construire une nouvelle voix de synthèse en utilisant un modèle TTS fondé sur l'apprentissage profond. L'assistant vocal de Sonos, le *Sonos Voice Control*, exploite le profil vocal de l'acteur de la série « Breaking Bad » Giancarlo Esposito. Esposito a passé environ 40 heures en studio à enregistrer des milliers de phrases et d'expressions qui ont ensuite été utilisées comme données d'entraînement pour le modèle vocal. Il a également lu des documents spécifiques à l'assistant vocal Sonos, même si ces phrases ne sont pas utilisées en 1:1. En plus, Sonos utilise les enregistrements vocaux de sa communauté d'utilisateurs *bêta* pour former le Sonos Voice Control.

La fabrique de la voix de synthèse et la protection de la voix humaine. Il convient dès lors de s'interroger sur la protection à accorder aux enregistrements de voix humaines exploités dans le cadre de la voix de synthèse consistant en un recueil de séries de mots, phonèmes, diphtonges prononcés avec différentes intonations par une ou plusieurs personnes dont la voix servira de modèle. La prestation réalisée par une personne pour procéder à ces enregistrements peut relever d'un contrat de travail et bénéficier de la rémunération afférente. A cet égard, l'usage est que la voix d'un comédien utilisée pour constituer la voix de synthèse des assistants vocaux le soit dans un cadre contractuel et rémunéré, celui-ci se faisant en général à la ligne ou à l'heure, comme cela peut l'être s'agissant du doublage ou de l'enregistrement d'une lecture de livres audio numériques. Au-delà, la question se pose de savoir si cet enregistrement peut également donner prise au droit voisin de l'artiste-interprète ou encore s'il peut être couvert par la protection conférée au titre des droits de la personnalité.

Protection au titre du droit voisin de l'artiste-interprète. Une première voie à explorer consiste à envisager la protection de la personne ayant réalisé ces enregistrements sonores au regard du droit voisin de l'artiste-interprète au sens de l'article L. 212-1 du CPI. Plusieurs conditions seront dès lors requises. La première résulte de la lettre du texte selon laquelle « *l'artiste-interprète ou exécutant est la personne qui représente, chante, récite, déclame, joue ou exécute de toute autre manière une œuvre littéraire ou artistique, un numéro de variété, de cirque ou de marionnettes* ». Il faut en déduire que « *l'artiste-interprète est la personne qui représente [...] de toute autre manière une œuvre de l'esprit* »⁹⁹ et qu' « *il n'y a pas*

⁹⁹ A. Lucas, A. Lucas-Schloetter et C. Bernault, *Traité de la propriété littéraire et artistique*, Lexis Nexis, 5e éd. 2017, n°1335.

d'interprétation sans œuvre»¹⁰⁰. Cette première condition peut alors poser difficulté s'agissant des enregistrements réalisés pour la constitution de la voix de synthèse d'un assistant vocal. Cela suppose en effet que ceux-ci soient issus de l'interprétation d'une œuvre de l'esprit, et non simplement des mots, de bribes de conversation ou des lectures de textes dépourvus de caractère créatif au sens de l'article L. 212-1 du CPI. La jurisprudence relative à l'image de la personne¹⁰¹ pourrait ici s'appliquer à celui qui ne fait que « prêter » sa voix à l'assistant vocal. Il convient en outre de s'interroger sur la nature de la prestation de la personne dont la voix est enregistrée. La qualification d'artiste-interprète suppose en effet que soit réalisée une interprétation traduisant une expression personnelle de l'interprète dans sa prestation, indépendamment du mérite de l'interprétation et de la notoriété de l'artiste, ce qui exclut de la protection toute personne dont le rôle demeure purement technique¹⁰². Dès lors, la fixation de la lecture d'une œuvre réalisée par un « diseur » (artiste interprète, journaliste) pourra être qualifiée d'interprétation au sens de l'article L. 212-1 du CPI à partir du moment où celle-ci n'est pas linéaire, et qu'elle révèle la personnalité de l'interprète. On rappellera par ailleurs que les artistes de doublage ou de voix off qui prêtent leur voix à un personnage principal bénéficient de la qualification d'artiste-interprète¹⁰³. La fixation d'une telle lecture devra être régie par un contrat d'artiste afin de prévoir les conditions de la cession de droits et la rémunération afférente qui, selon l'article L. 212-3 III du CPI, devra être appropriée et proportionnelle à la valeur économique réelle ou potentielle des droits cédés, compte tenu de la contribution de l'artiste-interprète à l'ensemble de l'œuvre et de toutes les autres circonstances de l'espèce, telles que les pratiques de marché ou l'exploitation réelle de la prestation, en cas d'exploitation monétisable.

Il convient de préciser ce qu'il en est dans l'hypothèse où l'enregistrement de cette interprétation est compilé avec d'autres extraits sonores afin de réaliser la voix de synthèse d'un assistant vocal. La première interrogation porte alors sur l'application du droit exclusif. A cet égard, l'arrêt InfoPaq a pu admettre que les parties d'une œuvre peuvent être « *protégées par le droit d'auteur dès lors qu'elles participent comme telles à l'originalité de l'œuvre entière* », après avoir rappelé que « *l'article 2, sous a), de la directive 2001/29 prévoit que les auteurs disposent du droit exclusif d'autoriser ou d'interdire la reproduction en tout ou en partie de leurs œuvres* »¹⁰⁴. La décision Pelhman confirme cette analyse en affirmant que le droit exclusif conféré par l'article 2, c) de la directive 2001/29 du 22 mai 2001 au producteur de phonogrammes

¹⁰⁰ A.E. Kahn, « Droits voisins du droit d'auteur. Droit des artistes-interprètes », *Jurisclasseur Propriété littéraire et artistique*, 2022, n°1.

¹⁰¹ V. notamment Paris, 18e ch. sect. D, 12 février 2008, n°C07/02721.

¹⁰² A.E. Kahn, préc., n°43 &s.

¹⁰³ Convention collective sur les droits des artistes interprètes de doublage, signée le 6 janvier 2005 et étendue le 12 mars 2005 - v. également Paris, ch. 2, 20 nov. 2015.

¹⁰⁴ CJUE, 16 juillet 2009, Infopaq International A/S contre Danske Dagblades Forening/Infopaq, aff. C-5/08, points 32 et 38. On relèvera que si la Cour de justice admet que « *les mots en tant que tels ne constituent pas des éléments sur lesquels portent la protection* » (point 46), elle retient que, « *compte tenu de l'exigence d'une interprétation large de la portée de la protection conférée par l'article 2 de la directive 2001/29, il ne saurait être exclu que certaines phrases isolées, ou même certains membres de phrases du texte concerné, soient aptes à transmettre au lecteur l'originalité d'une publication telle qu'un article de presse, en lui communiquant un élément qui est, en soi, l'expression de la création intellectuelle propre à l'auteur de cet article. De telles phrases ou de tels membres de phrase sont donc susceptibles de faire l'objet de la protection prévue à l'article 2, sous a), de ladite directive* » (point 47).

d'autoriser ou d'interdire la reproduction de son phonogramme « *lui permet de s'opposer à l'utilisation par un tiers d'un échantillon sonore, même très bref, de son phonogramme aux fins de l'inclusion de cet échantillon dans un autre phonogramme* ». La Cour de justice précise toutefois que cette disposition doit être interprétée à la lumière de la Charte des droits fondamentaux de l'Union européenne, ce dont il résulte que le jeu du droit de reproduction est conditionné par le fait que l'échantillon ne soit pas inclus dans un autre phonogramme « *sous une forme modifiée et non reconnaissable à l'écoute* »¹⁰⁵. Dans l'hypothèse où l'échantillon sonore utilisé par le créateur de la nouvelle œuvre est reconnaissable à l'écoute, la Cour admet que sa reproduction peut relever de l'exception de citation consacrée à l'article 5, 3, d. de la Directive 2001/29 lu ici encore à la lumière de la Charte, « *pour autant que ladite utilisation a pour objectif d'interagir avec l'œuvre sur laquelle l'échantillon a été prélevé* »¹⁰⁶. La formulation retenue par la Cour de justice peut conduire à admettre que telles limitations au droit exclusif supposent que l'extrait soit exploité pour constituer une œuvre nouvelle dans la mesure où elles se trouvent justifiées par la liberté des arts - liberté de création - protégée au titre de l'article 13 de la Charte¹⁰⁷. Dès lors, ces limites trouveraient difficilement à s'appliquer à des extraits d'enregistrements sonores compilés afin de constituer la voix de synthèse d'un assistant vocal, d'autant plus lorsqu'il s'agit d'un procédé à finalité exclusivement commerciale. Restera alors la seconde interrogation relative à l'assiette de la rémunération due au titre de l'exploitation d'extraits d'enregistrements d'interprétations d'une pluralité d'artistes-interprètes, qui pourrait ici relever d'une rémunération forfaitaire annuelle, sur le modèle des usages en matière de banques de sons proposées à l'achat.

Protection en qualité d'artiste de complément. A défaut d'exploitation protégée au titre du droit voisin, on pourra utilement envisager la qualification d'artiste de complément, défini par opposition à l'artiste-interprète, et pour laquelle deux critères sont traditionnellement retenus : l'absence de personnalité de la prestation, et le caractère interchangeable ou substituable de la prestation. Une décision relative à un enregistrement vocal réalisé par un comédien dans le cadre d'un contrat de travail pour une campagne publicitaire illustre le raisonnement retenu par la jurisprudence en ce domaine¹⁰⁸. Relevant le défaut d'interprétation d'une œuvre de l'esprit, les juges ont écarté la qualification d'artiste-interprète dès lors que la personnalité de l'artiste ne s'est pas exprimée dans le cadre de sa prestation dans la mesure où la voix ne se distinguait pas de celles des autres comédiens ayant prêté leur voix au personnage

¹⁰⁵ CJUE, 29 juillet 2019, Pelham GmbH, Moses Pelham, Martin Haas contre Ralf Hütter, Florian Schneider-Esleben, aff. C-476/17, point 39.

¹⁰⁶ CJUE, Pelham, préc., point 72.

¹⁰⁷ Sur l'exception de citation, v. toutefois CJUE 1er décembre 2011, Eva-Paria Painer, aff. C-145/10, point 136 ayant retenu que « *le point de savoir si la citation est faite dans le cadre d'une œuvre protégée par le droit d'auteur ou, au contraire, d'un objet protégé par un tel droit, est dépourvu de pertinence* ». Précisons toutefois que certains auteurs ont pu relever que la portée de cette décision ne saurait être trop largement entendue, la solution pouvant s'expliquer par les faits de l'espèce (la reproduction d'une photographie dans un article de presse qui constitue « une structure citante » même si celle-ci n'est pas qualifiée d'œuvre de l'esprit protégeable par le droit d'auteur) : M. Vivant et J.-M. Bruguière, *Droit d'auteur et droits voisins*, Dalloz, 2018, 4ème éd., n°657 ; également A. Lucas, A. Lucas-Schlotter, C. Bernault, *Traité de propriété littéraire et artistique*, Lexis Nexis, 5ème éd., 2017, n°465.

¹⁰⁸ Versailles, 9 oct. 2008 et les obs. de J.-M. Bruguière, *Propriétés intellectuelles* 2009, n° 31, p. 173.

lors d'enregistrements antérieurs. On rappellera en outre que la brièveté de la prestation ainsi que le caractère professionnel de la prestation demeurent à cet égard indifférents¹⁰⁹.

Perspectives. Différents projets sont aujourd'hui initiés, notamment pour des artistes de doublage, afin d'organiser une gestion collective volontaire s'agissant de l'exploitation de ces enregistrements vocaux. La constitution d'une base de données permettrait notamment d'identifier le « diseur » créancier du droit à rémunération et aux sociétés de gestion d'organiser cette gestion collective au regard des outils dont elles disposent. Compte tenu des progrès des techniques de traitement algorithmique de la voix, il apparaît nécessaire de soutenir de tels projets et d'inviter à une mobilisation des organismes de défense des intérêts de la profession afin de prévoir, encadrer et rémunérer l'emploi ou l'emprunt de voix d'artistes utilisées pour servir de matrice à ces systèmes.

Protection de la voix au titre des droits de la personnalité. Si la voix de synthèse se fonde sur des caractéristiques dérivées de personnes physiques, il convient d'envisager leur protection au titre des droits de la personnalité dans la mesure où la jurisprudence admet que la voix constitue comme un attribut de la personnalité, « *une sorte d'image sonore* »¹¹⁰. Comme cela a pu être relevé¹¹¹, la voix constitue un élément de la vie privée protégé sur le fondement de l'article 9 du Code civil en ce que, comme nous l'avons déjà souligné, elle permet d'identifier la personne, par le contenu du message qu'elle porte mais encore en tant que telle dès lors que l'on peut déduire de la voix de la personne son identité, âge, sexe, origine socioculturelle, état de santé voire son état émotionnel. La voix peut en outre s'analyser comme « *un attribut de la notoriété valorisable comme l'est l'image* »¹¹². Certaines décisions vont d'ailleurs jusqu'à admettre un droit autonome sur la voix¹¹³, que certains analysent comme un droit de la personnalité à part entière¹¹⁴. Suivant un mouvement de patrimonialisation similaire à celui connu s'agissant de l'exploitation de l'image, la personne se voit ainsi reconnaître un droit d'exploitation de sa voix, susceptible de commercialisation, comme en attestent les pratiques dans le secteur publicitaire ou s'agissant des sportifs et entraîneurs professionnels¹¹⁵. Par ailleurs, une personne peut en contester l'usage à des fins commerciales ou publicitaires dès lors qu'elle n'a

¹⁰⁹ V. not. Paris 31 mai 1996 et Cass. Civ. 1ère, 6 juillet 1999 pour une apparition de 2 secondes sur les 15 secondes du spot publicitaire.

¹¹⁰ TGI Paris, 19 mai 1982, *D.* 1983, jurisprudence p. 183 à propos de la diffusion non autorisée de la voix de Maria Callas - Paris, 22 janv. 2001, *D.* 2002, p. 2375, note A. Lepage.

¹¹¹ Par ex. TGI Paris, 3e ch., 1re sect., 27 sept. 2004, *CCE* 2004, comm. 153, obs. C. Caron - v. aux Etats-Unis, s'agissant de l'utilisation commerciale de la voix en méconnaissance d'un right of publicity, *Midler v. Ford Motor Co.*, 849 F. 2d 460 (9th Cir., 1988) : imitation dans une série de spots publicitaires de la voix de Bette Midler, cherchant à faire croire que celle-ci chantait elle-même et à tirer ainsi profit de sa célébrité ; *Waits v. Frito-Lay*, 978 F. 2d 1093 (9th Cir., 1992), cert. denied, 113 S. Ct. 1047 (1993) ; imitation de la voix de Tom Waits à des fins commerciales également (sound-alike commercial).

¹¹² J.-M. Bruguière et B. Gleize, *Droits de la personnalité*, Ellipses, 2015, n°183 - Également, C.-A. Maetz, *La notoriété. Essai sur l'appropriation d'une valeur économique*, PUAM, 2010, p. 251-252.

¹¹³ TGI Paris, 27 sept. 2004.

¹¹⁴ V. en ce sens Versailles, 9 octobre 2008 préc. - Adde, D. Huet-Weiller, « La protection juridique de la voix humaine », *RTD civ.* 1982, p. 497.

¹¹⁵ V. par ex. l'article L. 222-2-10-1 du Code du sport s'agissant des contrats conclus avec un sportif professionnel au titre de l'exploitation commerciale de son image, de son nom ou de sa voix - Adde, M. Serna, « La voix et le contrat : le contrat sur la voix », *CCC* 1999, chron. n°9.

pas entendu participer à ces activités ou invoquer un préjudice commercial qui serait constitué par un manque à gagner¹¹⁶. On peut ainsi en déduire qu'une personne pourra s'opposer à l'utilisation de son identité sonore pour générer une voix de synthèse, voire à ce que sa voix soit transformée ou sa prestation vocale prolongée au moyen d'un traitement algorithmique en l'absence d'autorisation de sa part.

Voix de synthèse : vers de nouveaux modèles ? On constate désormais le développement d'un nouveau marché de la voix de synthèse. Plusieurs entreprises proposent ce type de services, à l'image de Voxygen ou Acapella. Les usages sont divers, et permettent notamment de produire des films, documentaires ou émissions télévisuelles avec des acteurs ou des personnages historiques décédés à l'image du documentaire « Juger Pétain » réalisé à partir d'articles de presse et des minutes officielles du procès qui s'est tenu en 1945 et en utilisant un procédé de modélisation sonore développé par l'IRCAM ayant permis de reconstituer les voix et intonations des principaux protagonistes. Comme le souligne Nicolas Obin, chercheur au laboratoire de Sciences et technologies de la musique et du son de l'IRCAM, « *Nous sommes capables de reproduire artificiellement une voix humaine, à l'image de la voix du comédien André Dussolier qui peut désormais être produite par une machine et créer des phrases que le comédien n'a jamais prononcées, ou encore celle de Marilyn Monroe* ». Il précise que « *l'Ircam travaille sur les caractéristiques de la voix et est ainsi capable d'hybrider des voix, par exemple en mixant une voix humaine à celle d'un lion. Avec des logiciels évolués, il est possible aussi de modifier les attributs d'une voix : à partir d'un enregistrement neutre, il est envisageable d'ajouter des émotions à une voix, de la faire passer de la colère à la joie ou à la tristesse. L'enjeu de ces recherches consiste à créer des voix plus expressives, prochaine étape de la synthèse vocale...* »¹¹⁷. Les travaux conduits par ces équipes portent également sur l'adaptation de la voix de synthèse au regard de son contexte d'usage. Il s'agit alors d'adapter la voix de synthèse notamment pour ne pas générer, comme pour l'heure, une certaine monotonie lors de la lecture d'une œuvre. En outre, Nicolas Obin relève que si aujourd'hui « *la voix des assistants vocaux est souvent conçue sur un modèle de neutralité de façon à s'adapter à tous* », ces travaux permettent d'envisager d'autres modèles pour moduler l'identité vocale de l'assistant vocal en utilisant des intonations ou accents spécifiques. Au-delà, les variations de la voix pourraient favoriser de nouvelles formes d'interactions humain-machine en s'adaptant au contexte d'écoute voire aux émotions de l'utilisateur de l'assistant vocal, ce que permet désormais d'envisager le déploiement des chatbot reposant sur la technique de l'Affective computing.

¹¹⁶ *Traité Droit de la personnalité*, dir. J.-C. Saint Pau, LexisNexis, 2017, n°1233.

¹¹⁷ Interview réalisée par H. Guillaud, « Hypervoix 1 : concevoir les interfaces vocales », *Internetactu.net*, 24 mai 2019.

L'Affective computing

L'*Affective computing* vise à entraîner des ordinateurs en utilisant les émotions ou les affects humains. Il s'agit d'un domaine de recherche multidisciplinaire, dans lequel la sociologie, la psychologie, l'informatique, la physiologie, les mathématiques et la linguistique apportent une contribution importante. L'*Affective computing* comprend la procédure d'acquisition de signaux physiologiques et comportementaux, le traitement de ces signaux, l'extraction de *features* et la classification ou la prédiction. Ces caractéristiques extraites sont utilisées pour entraîner un modèle de données, qui peut ensuite classifier ou prédire des résultats dans divers domaines. Cette technologie a en effet recours à une diversité de techniques au service de trois finalités : détecter des émotions et « raisonner » sur le fondement de ces informations affectives, générer des expressions affectives nouvelles ; susciter des interactions humain-machine en se fondant « sur la notion de théorie de l'esprit, qui désigne l'aptitude cognitive d'un individu à attribuer des états mentaux inobservables (intention, désir, croyance, émotion) à soi-même ou à d'autres individus », autrement dit sur le fait que lorsqu' « un être humain perçoit une émotion chez autrui, il s'agit de perception subjective qui lui est propre »¹¹⁸.

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

Le déploiement de ces techniques qui permettent « de donner de multiples inflexions aux voix de synthèse, par exemple en leur permettant de suggérer, d'inciter, voire de faire culpabiliser celui qui les écoute »¹¹⁹ couplé au développement de l'*Affective computing* reposant sur le traitement des émotions de l'utilisateur de l'interface pour s'approcher du modèle de la conversation humaine soulèvent des questions relatives à la possible manipulation de l'utilisateur de l'assistant vocal. Cela conduit à s'interroger sur l'encadrement de ces outils et usages, tant en raison de leurs enjeux éthiques¹²⁰ que juridiques¹²¹. Sur ce dernier point, relevons que, pour l'heure, l'article 52 de la proposition de Règlement européen établissant des règles harmonisées concernant l'intelligence artificielle, en date du 21 avril 2021 (AI Act) impose aux fournisseurs de systèmes d'IA destinés à interagir avec des personnes physiques de veiller à ce que ceux-ci soient conçus et développés de manière à ce que les personnes physiques soient informées qu'elles interagissent avec de tels systèmes¹²², ce qui revient à n'imposer ici qu'une obligation de transparence¹²³.

¹¹⁸ Comité national pilote d'éthique du numérique, CNPEN, *Agents conversationnels : enjeux d'éthique*, avis n° 3, nov. 2021, p. 20.

¹¹⁹ N. Obin, Interview préc.

¹²⁰ Sur ce point, v. l'avis du CNPEN préc.

¹²¹ V. notamment J. Rochfeld et 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 Proposition de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) COM(2021), 206 final - v. également en ce sens les *Lignes directrices en matière d'éthique pour une IA digne de confiance* du Groupe d'experts de haut niveau nommé par la Commission européenne, publiées en 2019, selon lesquelles « Les systèmes d'IA ne devraient pas se présenter comme des êtres humains auprès des utilisateurs ; lorsqu'ils interagissent avec un système d'IA, les êtres humains ont le droit d'en être informés. Cet aspect implique que les systèmes d'IA doivent être identifiables en tant que tels » (point 78).

¹²³ Certains vont jusqu'à préconiser l'interdiction de ces systèmes de reconnaissance émotionnelle dont le Conseil de l'Europe souligne notamment l'impact sur l'intégrité psychologique des individus (Conseil de l'Europe, *Vers une régulation des systèmes d'IA. Perspectives internationales sur l'élaboration d'un cadre juridique fondé sur les normes du Conseil de l'Europe dans le*

Des possibilités de co-création impliquant la voix de l'utilisateur peuvent être envisagées à brève échéance, ce qui promet d'être un terreau plus fertile pour la voix de synthèse vocale par apprentissage automatique afin de personnaliser la relation humain-machine. Pour l'heure, en effet, l'apprentissage automatique gère difficilement les interfaces vocales personnalisées car elles ne peuvent relever de la seule exploitation des données de l'utilisateur. L'associer pourrait alors constituer une démarche déterminante afin de favoriser des profils d'assistants vocaux aux interfaces vocales plus singulières, plus inattendues¹²⁴, susceptibles de gommer à terme le décalage existant entre le niveau de perception de l'humain et celui de la machine. C'est pourquoi la création d'une voix de synthèse avec des quantités d'enregistrements de celle de l'utilisateur constitue une piste pour y parvenir. La modélisation des émotions et du raisonnement humain annoncent une configuration dans laquelle « *Demain, l'interface ce sera vous* »¹²⁵. De telles perspectives ne manquent pas de soulever une pluralité de questions : cet autre non-humain qui se construit dans le cadre d'une démarche anthropomorphique de l'utilisateur-utilisé, avec la voix de cet utilisateur, deviendra-t-il le miroir de son propre comportement verbal et ses propres limites verbales laissant ainsi entrevoir un rapide enfermement?¹²⁶ Est-ce que ce n'est pas l'évolution d'une logique d'interaction faisant courir le risque d'appauvrir le langage et la création si l'assistant vocal n'est pas assez performant? « *En effet, plus la similarité entre humain et machine augmentera plus l'écart au modèle pourra engendrer un sentiment de malaise voire d'appréhension* »¹²⁷. Au-delà, cela interroge sur le modèle de société à promouvoir et la volonté humaine de brider les modèles techniques pour éviter d'en arriver, à bref délai, à de telles extrémités. Auquel cas, de nombreux avatars verront le jour dans tous les univers créatifs avec la voix des utilisateurs étant entendu que de nombreuses start-ups proposent d'ores-et-déjà de créer son propre avatar vocal. Doit-on considérer alors que tous les utilisateurs seront à terme des interprètes¹²⁸, ayant vocation à remplacer les interprètes professionnels de manière à donner voix aux assistants virtuels et accentuer l'idée d'un clonage numérique de la personne ? Dans cette même perspective, sommes-nous en mesure d'accepter des avatars conçus pour l'éternité avec les voix des personnes décédées¹²⁹ ? La même technologie peut être utilisée pour les prothèses vocales, autrement dit le clonage de la voix¹³⁰. Ces possibilités de co-création de voix de synthèse personnifiées pour améliorer l'empathie de la relation humain-machine auront assurément un impact sur la

domaine des droits de l'homme, de la démocratie et de l'État de droit, DGI 2020(16), p. 26, point 20), sous réserve de certains usages spécifiques à finalités de recherche ou de protection de la santé (v. particulièrement CEPD-EDPS, *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence* (Artificial Intelligence Act), 18 juin 2021, point 35) ou de la démonstration que le système algorithmique permettrait de renforcer l'autonomie de la personne (v. CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, 7 avril 2022, A-2022-6, p. 14 et s.).

¹²⁴ H. Guillaud, préc., p.2.

¹²⁵ Expression de L. Julia, *Hypervoix*, FING, 2019, p. 33.

¹²⁶ FING, *Hypervoix. Promesses et illusions des nouvelles interfaces humain-machine*, janvier 2019

¹²⁷ Idem, p.34.

¹²⁸ F. Vallet, « Les droits de la voix (2/2) - Quelle parole pour nos systèmes ? », *LINC*, 27 juin 2019.

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

¹³⁰ V. sur ce point, entretien avec N. Obin, Livre blanc de la CNIL, préc., p.33 et s.

protection de la voix par un droit d'auteur qu'il convient de réinventer à l'aune des créations humains-machines, et un droit des artistes interprètes à adapter.

c. Le contenu de la réponse vocale : l'enjeu de la communication sous forme sonore d'une œuvre

La réponse de l'assistant vocal peut être de nature plurielle en fonction tant de la question posée par l'utilisateur de la technique de Text-to-Speech utilisée pour le fonctionnement de l'interface, allant de la réponse la plus simple et complètement contextuelle, à un discours plus enrichi en fonction du développement de la requête et de sa précision. A ce titre, dans le cadre du présent rapport, l'une des interrogations procède de la réponse apportée par l'assistant vocal qui pourrait consister en la communication d'une œuvre sous forme sonore par la technique de la synthèse vocale (Text-to-Speech), ce qui revient à déterminer si cette communication pourrait donner prise au droit de reproduction et au droit de communication au public.

Communication vocale et droit de reproduction. S'agissant du droit de reproduction, il convient de s'interroger sur le respect des conditions posées par l'article L. 122-3 du CPI aux termes duquel « *la reproduction consiste dans la fixation matérielle de l'œuvre par tous procédés qui permettent de la communiquer au public d'une manière indirecte* ». Celles-ci paraissent remplies en l'espèce. En effet, d'une part, il s'agit bien ici d'une fixation matérielle au sens du texte dès lors que la fixation peut être réalisée par tous procédés, et que « *la nature du support utilisé pour la fixation est rigoureusement indifférente* »¹³¹, ce qui comprend notamment la reproduction sur une mémoire informatique¹³² ou un espace de stockage distant¹³³. Tel sera le cas pour la communication d'une œuvre sous forme sonore générée par la technique de la synthèse vocale (c'est-à-dire la production artificielle de la voix humaine) qui suppose, pour ce faire, la mise en format numérique par exemple du texte, stocké soit sur le *hardware* (par exemple l'enceinte connectée) soit dans le cloud. D'autre part, l'œuvre a bien pour finalité d'être communiquée au public, ici à l'utilisateur de l'assistant vocal. En outre, le passage de l'écrit à l'oral demeure indifférent¹³⁴.

Il s'agit ensuite de vérifier si l'exception de reproduction provisoire pourrait trouver à s'appliquer. A cet égard, il convient de considérer les conditions posées par l'article L. 122-5, 6° du CPI lequel dispose que « *la reproduction provisoire présentant un caractère transitoire ou accessoire, lorsqu'elle est une partie intégrante et essentielle d'un procédé technique et qu'elle a pour unique objet de permettre l'utilisation licite de l'œuvre ou sa transmission entre tiers par la voie d'un réseau faisant appel à un*

¹³¹ A. Lucas, A. Lucas-Schloetter, C. Benaud, *Traité de propriété littéraire et artistique*, 5ème éd., 2017, n°259.

¹³² CJUE, 16 juillet 2009, Infopaq, Infopaq International A/S contre Danske Dagblades Forening, aff. C-5/08.

¹³³ TGI Paris, 14 août 1996 ou encore TGI Paris, 5 mai 1997.

¹³⁴ Cass. Civ. 1ère, 15 octobre 1985.

intermédiaire ; toutefois, cette reproduction provisoire qui ne peut porter que sur des œuvres autres que les logiciels et les bases de données ne doit pas avoir de valeur économique propre ». Quant au caractère provisoire tout d'abord, il en découle que la conservation « *ne peut excéder ce qui est nécessaire pour le bon fonctionnement du procédé technique* »¹³⁵ ; la copie doit donc être supprimée dès qu'elle a rempli sa fonction technique. Il revient alors d'en vérifier le caractère transitoire ou accessoire. Dans le cas de la reproduction de l'œuvre pour en réaliser une communication orale par la voie de synthèse, on pourra admettre le caractère accessoire de la copie dans la mesure où la finalité est de communiquer de façon sonore l'œuvre et non de la reproduire. Les conditions posées par la Cour de Justice dans l'arrêt NLA seront alors remplies car la reproduction n'aura « *ni existence ni finalité autonome* »¹³⁶. Il convient de relever que si, pour l'heure, la communication par la voix de synthèse est encore dans un état préliminaire (la lecture d'une œuvre consiste à ce jour à celle d'une application type podcast préenregistrée et par le biais de voix humaines), il est possible de considérer que les avancées techniques permettront dans un délai raisonnable la lecture d'une œuvre directement par l'assistant vocal à travers des techniques de Text-to-Speech. En revanche, il n'est pas encore possible de se prononcer sur les modalités de stockage du texte ou du speech généré après. On peut considérer par exemple que l'opérateur aura intérêt à stocker l'enregistrement de la voix de synthèse généré pour le rejouer à la suite de la requête d'autres utilisateurs pour éviter un nouveau traitement algorithmique. Dans ce cas, l'exception de copie provisoire ne trouverait pas à s'appliquer. Il conviendra enfin de vérifier l'absence de valeur économique propre de la copie, qui pourrait faire débat¹³⁷, notamment si l'on considère le flux de données en résultant.

Communication vocale et droit de communication au public. S'agissant du droit de représentation, l'article L. 122-2 du Code de la propriété intellectuelle dispose que « *la représentation consiste dans la communication de l'œuvre au public par un procédé quelconque* ». Celui-ci doit trouver à s'appliquer dans la mesure où la synthèse vocale permet une communication indirecte de l'œuvre à un public et que tous les éléments caractéristiques de l'œuvre sont communiqués par l'intermédiaire de ce procédé technique. En outre, conformément aux dispositions de l'article 3 de la Directive 2021/29, une telle communication consiste bien en la représentation de l'œuvre de manière à ce que chacun puisse y avoir accès de l'endroit et au moment qu'il choisit individuellement. Il en résulte que le titulaire de droit pourra s'opposer à la communication de l'œuvre par ce procédé en l'absence d'autorisation. Il convient en outre de relever que l'auteur pourra s'opposer à cette communication si la représentation vocale de l'œuvre porte atteinte au droit au respect de son intégrité, particulièrement en cas de représentation par la voix de synthèse de qualité médiocre.

¹³⁵ CJUE Infopaq, préc.

¹³⁶ CJUE 5 juin 2004, Public Relations Consultants Association Ltd contre Newspaper Licensing Agency Ltd e.a., aff. C-360/13.

¹³⁷ A cet égard, v. CJUE Infopaq, préc. écartant la signification économique indépendante « *pour autant, d'une part, que la mise en œuvre de ces actes ne permette pas de réaliser un bénéfice supplémentaire, allant au-delà de celui tiré de l'utilisation licite de l'œuvre protégée et que, d'autre part, les actes de reproduction provisoires n'aboutissent pas à une modification de l'œuvre* ».

B - L'assistant vocal comme nouvelle interface

Au-delà du système algorithmique, l'assistant vocal doit également se concevoir comme une interface permettant d'accéder de façon privilégiée, plus naturelle, à une pluralité de services proposés soit par le fournisseur de l'assistant vocal soit par des tiers, que l'utilisateur pourra choisir à partir d'un magasin d'applications. Une description de l'univers de l'assistant vocal (1) permet alors de mieux comprendre comment le cadre réglementaire s'en saisit progressivement en qualité de nouvelle interface, voire comme une forme particulière d'intermédiation (2).

1. Description de l'univers de l'assistant vocal

Assistant vocal et terminal. L'assistant vocal est un logiciel fonctionnant au moyen d'un dispositif physique, ou équipement terminal. La technologie d'un assistant vocal est représentée par un programme (*software*) qui est installé sur un dispositif ou terminal autonome (*hardware*). Plus précisément, un assistant vocal typique comporte deux couches : d'une part, la couche logiciel (qui comprend le *frontend*, c'est-à-dire le "visage" de l'application, constitué de l'interface utilisateur et de la technologie de reconnaissance vocale, et le *backend* qui exécute les services comme les algorithmes d'IA pour traiter une question, comprendre le langage naturel, repérer une réponse) ; d'autre part, la couche matériel (le terminal qui inclut les ordinateurs, les smartphones, les enceintes intelligentes, les autres appareils IoT, etc.). La technologie des assistants vocaux peut bénéficier et s'améliorer continuellement grâce aux dernières avancées en matière de logiciels et de matériel. De nombreuses solutions matérielles de ce type sont fermées (par exemple, les enceintes intelligentes) ; le logiciel peut également être propriétaire, comme le sont la plupart des assistants vocaux généralistes (Amazon Alexa, Apple Siri, Google Assistant).

Les applications vocales. Les applications vocales correspondent à des fonctionnalités permettant d'améliorer l'expérience utilisateur. Elles peuvent être activées manuellement ou déclenchées directement depuis l'enceinte avec la commande vocale. Ces applications sont conçues pour toutes les tâches qu'on souhaite pouvoir demander à l'assistant vocal d'exécuter, telles que fournir des actualités, des recettes de cuisine, jouer de la musique, trouver le meilleur moyen de transport pour un trajet donné, etc. Chaque plateforme a un nom différent pour ses applications vocales. Amazon utilise le terme *skills*,¹³⁸ Google et Siri optent pour le terme *applications*¹³⁹. Il est important de souligner que chaque plateforme possède ses propres applications intégrées, comme demander l'heure, la météo et de jouer un

¹³⁸ <https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html>

¹³⁹ <https://play.google.com/store> et <https://www.apple.com/fr/app-store/>

morceau de musique. Les applications vocales créées par les développeurs tiers peuvent être invoquées à l'aide d'une phrase spécifique ou, si l'assistant virtuel le permet, peuvent être invoquées implicitement, sans déclencher la phrase clé. Par exemple, dans l'invocation "*Hey Google, parle à <nom de l'application>*" il est explicitement indiqué quelle application est demandée. En revanche, l'invocation "*Hey Google, quel temps fait-il aujourd'hui à Paris ?*" sous-entend par le contexte de la demande quel service l'utilisateur veut, c'est-à-dire une application pour prédire la météo. Dans le cas d'une invocation implicite, l'assistant vocal pourra décider quelle application parmi celles installées et autorisées par l'utilisateur il utilisera pour répondre à sa requête, dans le cas où de multiples applications installées pourraient accomplir la même tâche. Dans le cas où l'assistant vocal n'a pas une application tierce installée et indiquée comme préférée par l'utilisateur pour une certaine tâche, il s'appuiera de préférence sur ses propres applications intégrées. L'association de comptes liés à une offre payante et l'utilisation d'applications vocales permettent idéalement une utilisation optimale et personnalisée des enceintes.

Les magasins d'applications et leurs enjeux. Un magasin d'applications est un type de plateforme de distribution numérique de logiciels informatiques appelés applications, souvent dans un contexte mobile. Les magasins d'applications se présentent généralement sous la forme d'une boutique en ligne, où les utilisateurs peuvent parcourir les différentes catégories d'applications, consulter des informations sur chaque application (telles que les évaluations des utilisateurs qui ont utilisé l'application) et acquérir l'application (l'achat de l'application, si nécessaire, ou le téléchargement d'applications proposées gratuitement). L'application sélectionnée est proposée en téléchargement automatique. Il s'en suit son installation. Toutes les applications, les mises à jour des applications, les packs d'applications, les achats *in-app* et les événements *in-app* soumis au magasin d'applications sont examinées par le propriétaire du magasin d'applications. Les principaux fabricants d'assistants vocaux disposent tous de leur propre magasin d'applications (par exemple, Alexa Skills, Apple App Store, Google Play, Microsoft Apps), où les utilisateurs peuvent sélectionner les applications qu'ils souhaitent installer et associer à leur profil. Cette opération permet à l'assistant vocal d'« appeler » et d'activer l'application parmi l'ensemble des applications disponibles dans le magasin. Souvent, cette opération inclut le fait que l'utilisateur doit se connecter avec son compte sur l'application pour permettre à l'assistant vocal d'utiliser cette application à la requête de l'utilisateur. Lorsqu'une nouvelle application est construite, il est nécessaire pour les développeurs d'utiliser les directives et les ressources fournies par le propriétaire du magasin d'applications pour que le processus d'examen de l'application avant sa publication dans le magasin se déroule bien. Cela pourrait conduire les opérateurs d'assistants vocaux à imposer leurs règles aux tiers, intensifiant de cette manière la complexité du processus de validation des applications tierces avant leur publication dans les magasins d'applications de ces mêmes opérateurs. Comme souligné dans l'enquête de la United States House of Representatives sur les marchés numériques, les acteurs du marché s'inquiètent alors de leurs difficultés à obtenir l'accès aux

fonctionnalités clés nécessaires à la création de leurs applications, telles que les commandes non traitées de l'utilisateur¹⁴⁰.

L'écosystème des assistants vocaux. Comme l'illustre la Figure 1 du rapport de la House of Representatives¹⁴¹, les assistants vocaux peuvent être vus comme une interface utilisateur qui permet des échanges entre des appareils informatiques par le biais d'une commande vocale d'une personne. Plus précisément, ces appareils informatiques peuvent prendre différentes formes telles que les enceintes intelligentes (Google Home, Amazon Echo Dot) ou les appareils mobiles et les ordinateurs. L'assistant vocal, connecté à ces appareils à travers le réseau, fournit une interface pour adresser des requêtes à ces appareils avec une commande vocale. A ces appareils, on peut ajouter aussi les appareils de domotique et IoT tels que des éclairages, des thermostats, des moniteurs de sécurité et même des appareils de cuisine qui sont connectés aux assistants vocaux à travers le *cloud*, un environnement de travail numérique relevant d'Internet pour stocker des données et exécuter des applications (c'est-à-dire les applications activées depuis le magasin d'applications).

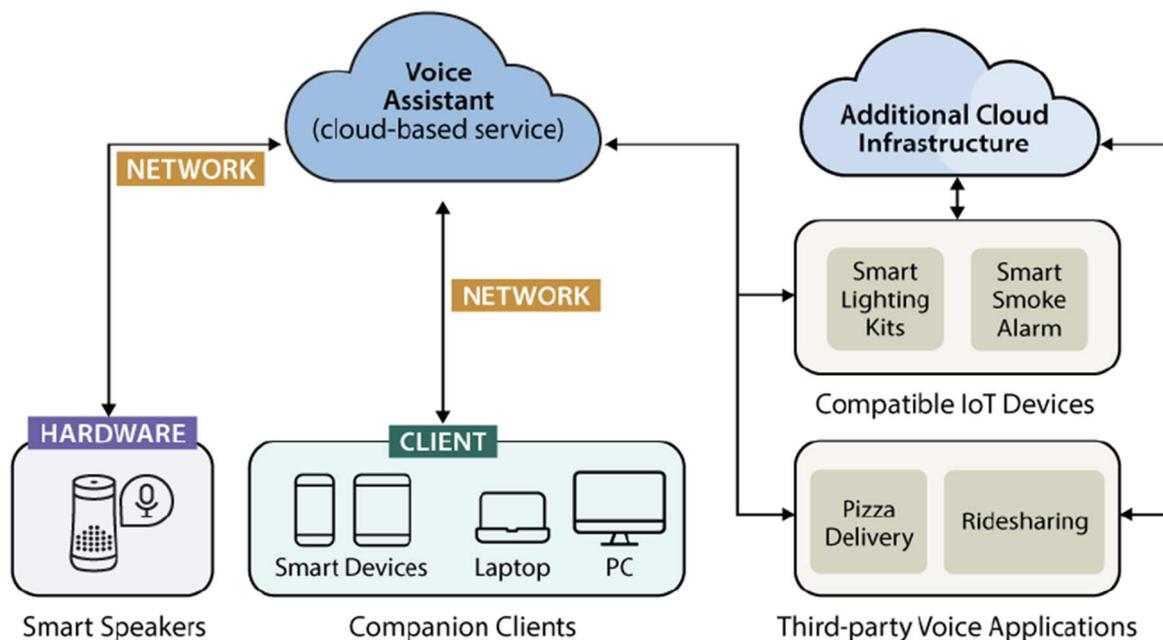


Figure 1 - Source : House of Representatives, *Investigation of competition in Digital Markets*, CP 117-8, Part 1, juillet 2022, p. 101

Deux dimensions se distinguent dans cet univers des assistants vocaux : (1) la dimension qui prend en compte les composantes matérielles et logicielles (*hardware*

¹⁴⁰ House of Representatives, *Investigation of the competition in digital markets*. Majority Staff Report and Recommendations. CP 117-8, Part 1, juillet 2022, p.

¹⁴¹ House of Representatives, enquête préc., 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/paperdigitalforensicapproachesforamazonalexalecosystem.pdf>

et *software*) autour de l'assistant vocal, et (2) la dimension qui prend en compte les applications et les magasins d'applications connectées à l'assistant vocal.

Pour ce qui concerne la dimension (1), la composante *hardware* inclut les enceintes connectées, les smartphones et les appareils de domotique, tandis que la composante *software* inclut l'assistant vocal lui-même en tant que logiciel (v. supra). L'assistant vocal permet d'émettre par la voix des requêtes aux composantes *hardware*. Comme souligné par le rapport de la House of Representatives : « *Le smartphone et le haut-parleur intelligent sont les deux principaux portails pour les assistants vocaux. Les acteurs du marché soulignent que les haut-parleurs intelligents représentent un hub ou une passerelle essentielle pour les maisons intelligentes et qu'ils favorisent l'adoption de l'assistant vocal. Les appareils IoT et les applications vocales peuvent être la propriété du fournisseur de la plateforme ou d'un tiers, si le fournisseur a mis en place des services permettant aux fabricants de créer des appareils compatibles avec l'assistant vocal* »¹⁴².

En ce qui concerne la dimension (2), dans ce contexte, l'assistant vocal représente une interface privilégiée pour accéder aux services propriétaires de son fournisseur. Cette interface privilégiée permet un accès préférentiel et une promotion « par défaut » des applications faisant partie du même univers que l'assistant vocal. Par exemple, l'assistant vocal Amazon Alexa privilégie l'écoute de la musique avec l'application (*skill*) qui fait partie de son propre univers, c'est-à-dire Amazon Music. Ces applications ne nécessitent pas une intégration explicite par l'utilisateur, différemment des applications tierces qui doivent être explicitement intégrées parmi les applications auxquelles l'assistant vocal peut accéder et qu'il peut lancer au moyen des commandes vocales.

Enjeux. Dans cet environnement, plusieurs enjeux peuvent résulter de l'interaction de l'assistant vocal avec son propre univers, tels que les effets de réseaux, le croisement de données d'usage des utilisateurs, l'autopréférence et le *lock-in strategy*.

Effet de réseau. Comme de nombreuses plateformes en ligne, le marché des assistants vocaux bénéficie d'effets de réseau. Plus précisément, plus il y a d'utilisateurs sur une plateforme, plus il y a d'appareils et d'applications tiers disponibles, ce qui attire davantage d'utilisateurs sur la plateforme. Ces effets de réseau pour les plateformes d'assistants vocaux sont amplifiés par les outils d'apprentissage automatique et l'Intelligence Artificielle. Les améliorations apportées aux algorithmes de traitement automatique du langage naturel et à l'IA devraient améliorer la qualité des assistants vocaux et contribuer à une adoption plus large.

Utilisation et croisement de données. Les données d'usage des utilisateurs incluent toutes les données sur des actions menées par l'utilisateur en interaction avec

¹⁴² House of Representative, Enquête préc., p. 102.

l'assistant vocal. Ces données permettent d'obtenir une amélioration dans la performance de l'assistant vocal, par exemple une amélioration de la compréhension de la requête de l'utilisateur. Plus précisément, la collecte et le croisement de ces données renforcent l'écosystème de certains opérateurs, permettant à ces acteurs d'améliorer leurs propres services, et en même temps d'obtenir une meilleure connaissance du profil de l'utilisateur. Ce profil devient donc plus complexe et n'inclut pas uniquement les données fournies explicitement par l'utilisateur mais aussi celles qui ont été collectées lors de ses interactions avec l'assistant vocal. Les inférences sur ces données grâce à la mise en place d'algorithmes de croisement des données permettent d'acquérir une connaissance approfondie de l'utilisateur et de ses préférences. Cela ouvre aux enjeux de protection des données personnelles, déjà évoqués ci-dessus. Le résultat est la proposition d'une offre plus attractive en termes de réponses et services fournis à l'utilisateur ainsi qu'une meilleure « position sur le marché ». Notamment, l'interaction avec l'utilisateur devient plus naturelle car l'assistant vocal a appris à connaître l'utilisateur par ses interactions passées.¹⁴³ En outre, le coût nécessaire pour stocker ces données peut être moins élevé quand l'opérateur de l'assistant vocal peut s'appuyer sur d'autres composantes de son univers, comme le cloud, permettant ainsi à l'opérateur de proposer des offres pour le stockage des données et de services à moindre coût. Pour résumer, l'utilisation des données d'usage et du profil utilisateur risque de favoriser la position des opérateurs de l'assistant vocal sur le marché.

Autopréférence et autopromotion. Les opérateurs des assistants vocaux peuvent se servir de toutes ces données afin de promouvoir d'autres services disponibles dans leur univers, par exemple une application de l'opérateur dans le magasin d'applications. Cela déclenche des enjeux d'autopréférence sur les deux dimensions mentionnées ci-dessous. Ainsi, l'enceinte de l'opérateur d'assistant vocal est utilisée de préférence pour interagir avec l'assistant vocal et les applications qui seront privilégiées pour répondre aux requêtes de l'utilisateur seront celles du même opérateur (par exemple, l'enceinte Amazon Echo Dot emploie l'assistant vocal Amazon Alexa pour écouter de la musique avec l'application Amazon Music). L'absence d'écran dans l'interaction avec les assistants vocaux ainsi que les défis liés à la réponse unique sur ces systèmes peuvent conduire à favoriser les services de ces opérateurs par défaut.

2. L'assistant vocal saisi par les textes comme une nouvelle forme d'intermédiation

Comme décrit précédemment, l'assistant vocal ou virtuel se conçoit comme une nouvelle interface au centre de la relation entre le fournisseur d'assistant vocal et l'utilisateur de celui-ci, permettant à ce dernier d'accéder à des services connexes,

¹⁴³ En plus, l'utilisateur améliore sa capacité à interagir avec l'assistant vocal.

proposés soit par le fournisseur d'assistant vocal soit par des entreprises tierces. Il constitue en outre une fonctionnalité supplémentaire d'accès à un écosystème, voire peut devenir le point d'entrée permettant d'accéder à un ensemble de services numériques proposés par son fournisseur. Par conséquent, l'assistant vocal ou virtuel caractérise une nouvelle forme d'intermédiation pour laquelle il convient de penser un cadre juridique pertinent. Tel est l'objet de récentes réformes du droit de l'Union européenne s'agissant des marchés et services numériques. Plusieurs textes visent désormais explicitement les assistants vocaux ou virtuels à l'aune de cette nouvelle forme d'intermédiation.

Introduction des assistants vocaux en droit de l'UE. Le premier texte à avoir expressément visé les assistants vocaux a été le Règlement du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne dit Platform to business¹⁴⁴ ; cette précision avait été apportée à la toute fin des négociations du texte afin d'appréhender une forme d'interface qui commençait à se déployer en Europe, et particulièrement en Allemagne. Son considérant 11 précise que les services d'intermédiation en ligne couverts par le présent règlement englobent, « *à titre d'exemple, les places de marché pour le commerce électronique, y compris les places collaboratives où les entreprises utilisatrices sont présentes, les services d'applications logicielles en ligne, tels que les boutiques d'applications, et les services de réseaux sociaux en ligne, quelle que soit la technologie utilisée pour fournir de tels services. En ce sens, les services d'intermédiation en ligne pourraient aussi être fournis par la technologie d'assistance vocale* ». Quelques mois après, c'est le droit de la consommation de l'Union européenne qui s'en saisissait à l'occasion de la réforme transversale de l'acquis réalisée par la Directive 2019/2161 dite Directive Omnibus¹⁴⁵, particulièrement pour préciser que la forme imposée pour respecter les obligations d'informations prescrites par le texte devra être revue lorsque le contrat est conclu par l'intermédiaire d'un « assistant vocal d'achat » dès lors qu'il s'agit d'une commande vocale¹⁴⁶. Son considérant 41 énonce ainsi que l'article 8, paragraphe 4 de la directive 2011/83/UE concerne les contrats à distance « *comme par téléphone, par l'intermédiaire d'un assistant d'achat vocal ou par SMS* »¹⁴⁷.

Nouvelle forme d'intermédiation et fonctionnement des marchés numériques : le DMA. Le Règlement sur les marchés numériques (ou Digital Market Act)¹⁴⁸ intègre dans son champ d'application les assistants virtuels. L'article 2.12 du règlement les

¹⁴⁴ Règlement (UE) 2019/1150 du Parlement européen et du Conseil du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne.

¹⁴⁵ Directive 2019/2162 du Parlement européen et du Conseil du 27 novembre 2019 en ce qui concerne une meilleure application et une modernisation des règles de l'Union en matière de protection des consommateurs.

¹⁴⁶ V. déjà sur cette question, K. Sein, « Concluding Consumer Contracts via Smart Assistants: Mission Impossible Under European Consumer Law? », *EuCML* 2018, 179.

¹⁴⁷ V. également Communication de la Commission, *Orientations concernant l'interprétation et l'application de la directive 2011/83/UE du Parlement européen et du Conseil relative aux droits des consommateurs* (2021/C 525/01), point 3.2.2.2.

¹⁴⁸ Règlement (UE) 2022/1925 du Parlement européen et du Conseil du 14 septembre 2022 relatif aux marchés contestables et équitables dans le secteur numérique et modifiant les directives (UE) 2019/1937 et (UE) 2020/1828 (Règlement sur les marchés numériques).

définit comme « *un logiciel qui peut traiter des demandes, des tâches ou des questions, notamment celles fondées sur des données d'entrée sonores, visuelles ou écrites, de gestes ou de mouvements, et qui, sur la base de ces demandes, tâches ou questions, donne accès à d'autres services ou contrôle des appareils connectés physiques* ». Le texte permet alors de saisir les assistants vocaux sous leurs multiples facettes, à savoir à la fois comme une nouvelle interface utilisateur mais aussi, et surtout, comme une porte d'entrée vers de multiples marchés.

Rappelons que la démarche portée par le DMA se veut globale et prospective, à mi-chemin entre le droit des pratiques restrictives de concurrence et l'approche par la régulation¹⁴⁹. Le texte s'applique aux plus grands contrôleurs d'accès qui jouissent d'une position hégémonique et stable leur conférant un pouvoir économique considérable. Cette réforme part d'un constat : ces opérateurs relient de nombreuses entreprises utilisatrices et de nombreux utilisateurs finaux, ce dont ils tirent avantages pour accroître leur pouvoir sur l'écosystème dont ils sont l'épicentre notamment grâce à l'accès à de vastes quantités de données. Il peut en résulter de graves déséquilibres et constituer la source de pratiques et conditions déloyales à l'égard des entreprises utilisatrices de leur service ainsi que des utilisateurs finaux. Dès lors, le texte consacre à la charge de ces opérateurs des obligations *ex ante* afin d'éviter qu'ils ne verrouillent le marché dont ils constituent le point d'accès, en empêchant l'entrée de concurrents ou en imposant leurs règles de fonctionnement dans leur intérêt (ex. interdiction des pratiques dites d'autopréférence consistant à favoriser ses propres services, ou encore de la combinaison et du croisement de données entre services). D'autres obligations visent quant à elles à déverrouiller le marché pour en favoriser l'accès à des tiers (ex. principe de non-discrimination, interopérabilité, accès et portabilité des données)¹⁵⁰.

Avec l'adoption du DMA, l'assistant virtuel est inclus dans la définition des services de plateforme essentiels, alors que ce n'était pas le cas lors des premières phases de négociation du texte. Cet ajout, proposé par la Commission du marché intérieur et de la protection des consommateurs du Parlement européen puis validé par le Conseil, peut s'expliquer par la conduite, au cours des négociations, de l'enquête sectorielle réalisée par la Commission européenne sur l'Internet des objets. Cette enquête a en effet mis en exergue les enjeux et importantes préoccupations des acteurs quant à l'incidence négative des pratiques des principaux fournisseurs d'assistants vocaux sur la concurrence, l'innovation et le choix des consommateurs dans le secteur de l'IdO ; ce à quoi s'ajoutait l'annonce d'une enquête lancée par la Commission européenne à propos de Google Assistant compte tenu de son installation par défaut sur les appareils fonctionnant sous Android¹⁵¹.

¹⁴⁹ En ce sens, v. J.-C. Roda, « Un an de droit de la concurrence dans l'univers numérique », *CCE* 2022, étude n°12, n°2 et M.-A. Frison-Roche, *Droit de la concurrence*, Dalloz, 2022, n°665 &s. - également S. Abiteboul et J. Cattan, *Nous sommes les réseaux sociaux*, Odile Jacob, 2022, p. 182 &s.

¹⁵⁰ Sur ces différentes obligations, v. notamment M. Debroux, « Nature, objectifs et régime des obligations imposées aux contrôleurs d'accès : Des obligations *ex ante* et *per se*, vraiment ? », *Concurrences* n°3-2022, pp. 66-72. Adde, M. Le Roy, *La loyauté des plateformes à l'égard des consommateurs*, Dalloz bibliothèque des thèses, 2023, n°274 &s.

¹⁵¹ <https://www.competitionpolicyinternational.com/eu-probes-googles-voice-assistant/>

Si les assistants virtuels entrent dans le champ d'application du DMA au titre des services de plateforme essentiels, il conviendra néanmoins d'envisager la qualification de contrôleur d'accès pour soumettre ces services de plateforme essentiels aux obligations posées par le texte. Rappelons en effet que la logique du DMA est de s'intéresser à ces grandes entreprises en ce qu'elles fournissent des services. Une fois retenue cette qualification, la seconde étape supposera de savoir quels sont les services de plateforme essentiels fournis par ces entreprises. A ce titre, sont considérés comme contrôleurs d'accès, ceux qui remplissent les critères quantitatifs suivants. Il s'agit d'entreprises ayant un chiffre d'affaires ou une valorisation boursière très élevé : 7,5 milliards d'euros au moins de chiffre d'affaires annuel en Europe dans les trois dernières années ou 75 milliards d'euros ou plus de capitalisation boursière durant la dernière année. Dans la mesure où 90% des entreprises du numérique (10 000 environ) sont considérées comme des PME dans l'impossibilité de capter le marché, rétablir l'accès au marché suppose de poser des obligations aux 10% restantes et de considérer que les 10 000 entreprises en question sont utilisatrices des services fournis par les contrôleurs d'accès. Dans le texte, les services entrant dans le champ d'application sont désignés comme des « services de plateforme essentiels » et sont au nombre de 10. Il s'agit des services d'intermédiation (comme les places de marché, les boutiques d'applications) ; les moteurs de recherche ; les réseaux sociaux ; les plateformes de partage de vidéos ; les messageries en ligne ; les systèmes d'exploitation (dont les télévisions connectées) ; les services en nuage (*cloud*) ; les services publicitaires (tels les réseaux ou les échanges publicitaires) ; les navigateurs web et les assistants virtuels. Ces 10 services sont problématiques au sens du DMA, et se trouvent alors soumis aux obligations posées par le texte, s'ils sont fournis au moins dans trois pays européens ; s'ils enregistrent plus de 45 millions d'européens par mois et s'ils utilisent les services de 10 000 professionnels par an pendant les trois dernières années. Pour résumer, une fois les entreprises désignées comme contrôleur d'accès, la Commission devra lister tous les services qui remplissent ces conditions, qualifiés alors de services de plateforme essentiels.

Le texte du DMA pose pour les assistants virtuels, d'autres définitions, et notamment précise ce que l'on entend par nombre d'utilisateurs finaux uniques, à savoir ceux, qui « *au moins une fois pendant le mois, ont interagi avec l'assistant virtuel de quelque manière que ce soit, par exemple en l'activant, en posant une question, en accédant à un service par une commande ou en contrôlant un dispositif de maison intelligente* ». Au titre du DMA, le nombre d'entreprises utilisatrices vise le « *nombre de développeurs uniques qui, au cours de l'année, ont proposé au moins une application logicielle d'assistant virtuel ou une fonctionnalité permettant de rendre une application logicielle existante accessible par l'intermédiaire de l'assistant virtuel* ». Dès lors, la question de l'application du DMA aux services de plateforme essentiels reste suspendue au respect des seuils par les différents assistants virtuels présents sur le marché. Cependant, l'intégration des assistants vocaux au cours des négociations du texte permet, à tout le moins, de percevoir le potentiel de ces acteurs en devenir, et la

possibilité de les saisir à l'aide des critères qualitatifs posés par le texte à l'article 3.8 dans la mesure où il est pour l'heure difficile de savoir s'ils remplissent les seuils quantitatifs en l'absence de données chiffrées. Si l'entreprise visée n'atteint pas les seuils visés au paragraphe 2 de l'article 3, la Commission peut en effet, en application de l'article 3.8 du DMA, désigner comme contrôleur d'accès toute entreprise qui 1° a un poids important sur le marché intérieur ; 2° fournit un service de plateforme essentiel qui constitue un point d'accès majeur permettant aux entreprises utilisatrices d'atteindre leurs utilisateurs finaux; et 3° jouit d'une position solide et durable, dans ses activités, ou jouira, selon toute probabilité, d'une telle position dans un avenir proche. A cette fin, « *la Commission tient compte de tout ou partie des éléments ci-après, pour autant qu'ils soient pertinents pour l'entreprise considérée fournissant des services de plateforme essentiels* :

- a) *la taille, y compris le chiffre d'affaires et la capitalisation boursière, les activités et la position de ladite entreprise ;*
- b) *le nombre d'entreprises utilisatrices qui font appel au service de plateforme essentiel pour atteindre des utilisateurs finaux et le nombre d'utilisateurs finaux ;*
- c) *les effets de réseau et les avantages tirés des données, en particulier en ce qui concerne l'accès aux données à caractère personnel et non personnel et la collecte de ces données par ladite entreprise, ou les capacités d'analyse de cette dernière ;*
- d) *tout effet d'échelle et de gamme dont bénéficie l'entreprise, y compris en ce qui concerne les données et, le cas échéant, ses activités en dehors de l'Union ;*
- e) *la captivité des entreprises utilisatrices ou des utilisateurs finaux, y compris les coûts de changement et les biais comportementaux qui réduisent la capacité des entreprises utilisatrices et des utilisateurs finaux à changer de fournisseur ou à opter pour un multihébergement ;*
- f) *une structure d'entreprise conglomerale ou l'intégration verticale de cette entreprise, permettant par exemple à celle-ci de pratiquer des subventions croisées, de combiner des données provenant de différentes sources ou de tirer parti de sa position ; ou*
- g) *d'autres caractéristiques structurelles des entreprises ou des services ».*

Il ressort donc de l'article 3.8 du DMA que l'approche écosystémique doit guider la Commission dans le cadre de la qualification des contrôleurs d'accès, ce qui est particulièrement important pour saisir l'activité des services numériques visés par le texte¹⁵². Par ailleurs, il convient de relever que, si la fourniture de l'assistant vocal n'est pas considérée comme un service de plateforme essentiel, l'entreprise qualifiée de contrôleur d'accès pourra voir ses activités encadrées par le texte conformément à cette même approche systémique. En effet, pour certaines obligations prescrites, « *le DMA ne raisonne pas par service mais bien par écosystème de l'entreprise concernée* ». Par exemple, conformément à l'article 5.2, « *l'entreprise contrôleur*

¹⁵² Sur la nécessité de réguler les entreprises et leurs écosystèmes, v. J. Toledano, *Gafa. Reprenons le pouvoir !*, Odile Jacob 2020, p. 128 &s.

*d'accès ne pourra faire une utilisation croisée des données acquises dans le cadre de ses différents services »*¹⁵³.

Au-delà, il peut être observé que la logique du texte permet d'identifier des obligations qui seraient applicables en général aux assistants vocaux si l'entreprise fournissant l'assistant vocal est désignée comme contrôleur d'accès. Compte tenu de la diversité des modèles d'affaires des assistants vocaux, il paraît toutefois nécessaire de préconiser des obligations spécifiques pour tel ou tel d'entre eux, en fonction de leur position sur le marché et de leurs manœuvres pour en contourner les règles. L'élaboration d'actes délégués prévue par le DMA pour permettre à la Commission d'étendre les obligations posées par le texte selon le dispositif envisagé à l'article 8, ne pourra pas créer cette asymétrie car ils sont conçus comme des actes d'application des règlements, qui ne sauraient aller au-delà de la lettre du texte. Si l'approche par la régulation *ex ante* est portée par le texte, elle reste dès lors symétrique. Ceci explique pourquoi certains plébiscitent une évolution du texte à venir vers une logique plus proche du droit de la régulation permettant d'adapter ces obligations « sur mesure » à chaque opérateur de façon plus efficace et réactive pour éviter que le droit de l'Union ne reste au milieu du gué¹⁵⁴.

Nouvelle forme d'intermédiation et partage de la donnée. La proposition de Règlement dit Data Act¹⁵⁵, en cours de négociation, vise également les assistants vocaux en reprenant, quasiment à l'identique, la définition du DMA (« *des logiciels capables de traiter des demandes, des tâches ou des questions, notamment à partir de données d'entrée sonores ou écrites, ou de gestes ou de mouvements, et qui, sur la base de ces demandes, tâches ou questions, permettent d'accéder à leurs propres services et à des services tiers, ou contrôlent leurs appareils et des appareils tiers* »). S'inscrivant dans la stratégie européenne en matière de données aux côtés du Règlement sur la gouvernance des données qui a défini un cadre de gouvernance pour le commerce des données industrielles, ce texte doit contribuer à ouvrir un marché pour les données produites par les appareils connectés. Son modèle de législation horizontale affectera, comme le DMA, des modèles commerciaux, des produits et des services hétérogènes différents. Dans ce contexte, le Data Act souligne le rôle essentiel des assistants vocaux dans le cadre de l'Internet des objets, et l'intérêt d'un partage des données au profit du consommateur ou entre entreprises, dans le champ d'application duquel tomberont les assistants vocaux (article 7). L'objectif : favoriser le développement d'une « *offre concurrentielle de services après-vente, une innovation plus large fondée sur les données et le développement de produits ou de services non liés à ceux que l'acheteur a initialement achetés ou auxquels il s'est abonné* ».

¹⁵³ D. Bosco, « A propos du concept de contrôleur d'accès », *Concurrence*, n°3-2022, p. 61, spéc. p. 63.

¹⁵⁴ K. Favro et C. Zolynski, « DSA-DMA : l'Europe au milieu du gué », *Dalloz IP IT* 2021, pp. 217-223.

¹⁵⁵ Proposition de Règlement du Parlement européen et du Conseil fixant des règles harmonisées pour l'équité de l'accès aux données et de l'utilisation des données, COM(2022) 68 final, 23 février 2022.

Proposition de Règlement Data Act, considérant 22

Considérant 22 : « *Les assistants virtuels jouent un rôle croissant dans la dématérialisation de l'environnement des consommateurs et servent d'interface facile à utiliser pour jouer des contenus, obtenir des informations ou activer des objets physiques connectés à l'internet des objets. Ils peuvent servir de portail unique dans un environnement domestique intelligent, par exemple, et enregistrer des quantités importantes de données utiles sur la manière dont les utilisateurs interagissent avec les produits connectés à l'internet des objets, dont ceux fabriqués par d'autres parties, et ils peuvent remplacer l'utilisation d'interfaces fournies par le fabricant telles que des écrans tactiles ou des applications pour smartphones. L'utilisateur pourrait souhaiter mettre ces données à la disposition de fabricants tiers et ainsi permettre l'avènement de services domotiques nouveaux. Ces assistants virtuels devraient relever du droit d'accès aux données prévu par le présent règlement également en ce qui concerne, d'une part, les données enregistrées avant l'activation de l'assistant virtuel par le mot déclencheur et, d'autre part, les données générées lorsqu'un utilisateur interagit avec un produit par l'intermédiaire d'un assistant virtuel fourni par une entité autre que le fabricant du produit. Toutefois, seules les données provenant de l'interaction entre l'utilisateur et le produit par l'intermédiaire de l'assistant virtuel relèvent du champ d'application du présent règlement. Les données produites par l'assistant virtuel qui sont sans rapport avec l'utilisation d'un produit ne sont pas l'objet du présent règlement* ».

L'assistant vocal, une nouvelle forme d'intermédiation au sens du DSA ? Le Digital services Act (ou Règlement sur les services numériques)¹⁵⁶ vise à encadrer les pratiques des services numériques pour en corriger les dysfonctionnements en responsabilisant les opérateurs de services numériques, sur le modèle de la régulation financière. L'objectif du texte est alors de protéger les droits fondamentaux, d'assurer un environnement sécurisé par la mise en place d'une nouvelle gouvernance, et répondre à des problématiques évoluant au jour le jour. A cette fin, l'Union européenne a fait le choix de mettre en place une régulation transversale et cohérente des services numériques. Le texte trouve à s'appliquer à une pluralité d'acteurs dont le modèle d'affaires et les problématiques sous-jacentes sont pourtant éminemment hétérogènes. Le DSA promeut ainsi une régulation par la transparence des services d'intermédiation. L'approche horizontale visée par le DMA n'est qu'imparfaitement restituée au sein du DSA qui se contente, quant à lui, d'une double approche verticale visant les plateformes en lignes et moteurs de recherche.

Il convient par ailleurs de préciser que le DSA ne saisit pas la même fonction s'agissant de l'intermédiation. Il s'applique à tous les intermédiaires en ligne qui offrent leurs services (biens, contenus ou services) sur le marché intérieur et, plus spécifiquement, à l'activité d'intermédiation *per se* de la plateforme (par ex. son

¹⁵⁶ Règlement 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (Règlement sur les services numériques).

activité de modération de contenus de tiers ou la recommandation du contenu de tiers par l'usage d'un algorithme de recommandation). C'est une intermédiation d'une autre nature que saisit le DMA, qui a lui vocation à appréhender l'activité de porte d'entrée sur un marché de l'opérateur, activité qui lui permet de définir les règles d'organisation de celui-ci. L'assistant vocal ne paraît pas répondre à la qualification d'intermédiaire au sens du DSA à l'égard duquel il demeure une simple interface, ce qui explique qu'il ne soit pas visé formellement par le texte.

Néanmoins, l'esprit de ce texte ouvre le champ des possibles au regard de l'intention et des obligations qui y sont formulées. S'agissant de l'intention, le considérant 12 précise qu' « *afin d'atteindre l'objectif consistant à garantir un environnement en ligne sûr, prévisible et fiable, il convient, aux fins du présent règlement, que la notion de « contenu illicite » corresponde de manière générale aux règles en vigueur dans l'environnement hors ligne. Il convient, en particulier, de donner une définition large de la notion de « contenu illicite » de façon à ce qu'elle couvre les informations relatives aux contenus, produits, services et activités illégaux. En particulier, cette notion devrait être comprise comme se référant à des informations, quelle que soit leur forme, qui, en vertu du droit applicable, sont soit elles-mêmes illicites, comme les discours haineux illégaux ou les contenus à caractère terroriste et les contenus discriminatoires illégaux, soit rendues illicites par les règles applicables en raison du fait qu'elles se rapportent à des activités illégales. Il peut s'agir, par exemple, du partage d'images représentant des abus sexuels commis sur des enfants, du partage illégal d'images privées sans consentement, du harcèlement en ligne, de la vente de produits non conformes ou contrefaits, de la vente de produits ou de la fourniture de services en violation du droit en matière de protection des consommateurs, de l'utilisation non autorisée de matériel protégé par le droit d'auteur, de l'offre illégale de services de logement ou de la vente illégale d'animaux vivants. (...) Il importe peu à cet égard que l'illégalité de l'information ou de l'activité procède du droit de l'Union ou du droit national conforme au droit de l'Union et il est indifférent de connaître la nature ou l'objet précis du droit en question* ». Le fonctionnement de l'assistant vocal en qualité d'intermédiaire entre les entreprises utilisatrices tierces et les utilisateurs pourrait être qualifié de pratiques régies par le texte visant par exemples le paramétrage (article 25) et autres interfaces trompeuses ou la publicité trompeuse. Néanmoins, si ces pratiques identifiées dans le fonctionnement des assistants vocaux ne sont pas saisies par la qualification de services de plateformes en ligne, à savoir par le DSA, il faudra les saisir par le marché.

L'enjeu particulier : l'agent conversationnel et régulation de l'IA (AI Act). Dans la mesure où l'agent conversationnel repose sur un système algorithmique d'apprentissage automatique, il pourra être saisi par le cadre juridique en construction s'agissant de la régulation de l'Intelligence artificielle. A cet égard, la proposition de Règlement établissant des règles harmonisées concernant l'intelligence artificielle dit

Règlement IA¹⁵⁷ ne qualifie *per se* les agents conversationnels ni comme des systèmes à très hauts risques - frappés par un principe d'interdiction de déploiement dans le marché intérieur - ni comme un système à haut risque - qui se trouve encadré par les règles de mise en conformité consacrées par le texte. Dans son communiqué de presse en date du 21 avril 2022, la Commission européenne énonce de façon univoque la classification des agents conversationnels : il s'agit d'applications ou de systèmes d'intelligence artificielle ne présentant qu'un risque limité pour les droits fondamentaux des personnes¹⁵⁸. Le texte n'impose alors que des obligations de transparence au fournisseur du système d'IA. Ainsi, son article 52 prévoit que « *les fournisseurs veillent à ce que les systèmes d'IA destinés à interagir avec des personnes physiques soient conçus et développés de manière à ce que les personnes physiques soient informées qu'elles interagissent avec un système d'IA, sauf si cela ressort clairement des circonstances et du contexte d'utilisation* ». Il s'agit alors de prémunir l'utilisateur final d'une possible confusion de statut lors de son interaction avec la machine¹⁵⁹. Notons que le Boltering Online Transparency Act of California Senate Bill 1001 (dit « Bot Act »)¹⁶⁰ prévoit également qu' il « *est illégal pour toute personne d'utiliser un bot pour communiquer ou interagir avec une autre personne en Californie en ligne, avec l'intention de tromper l'autre personne sur son identité artificielle dans le but de tromper sciemment la personne sur le contenu de la communication afin d'inciter à l'achat ou à la vente de biens ou de services dans une transaction commerciale ou d'influencer un vote dans une élection* ». Quant au paragraphe 2 de l'article 52 du Règlement IA, celui-ci vise plus précisément les systèmes de reconnaissance émotionnelle, précisant à cet égard que « *les utilisateurs d'un système de reconnaissance des émotions ou d'un système de catégorisation biométrique informent du fonctionnement du système les personnes physiques qui y sont exposées. Cette obligation ne s'applique pas aux systèmes d'IA de catégorisation biométrique dont la loi autorise l'utilisation à des fins de prévention et de détection des infractions pénales et d'enquêtes en la matière* ». Il ressort donc de ces dispositions, toujours en cours de négociation à ce jour, que c'est ici la nature particulière de l'interface humain-machine ainsi que les enjeux de manipulation qui peuvent en résulter à l'égard de l'utilisateur final qui justifient la consécration d'une - simple - obligation de transparence à l'égard du fournisseur du système d'IA¹⁶¹.

Les textes récemment adoptés ou en cours de négociation pourraient offrir de nouvelles perspectives qui permettraient d'appréhender la fonction d'intermédiation propres aux assistants vocaux. Ces réformes sont particulièrement justifiées en raison de la place centrale qu'occupe désormais cette interface au sein des écosystèmes, en qualité de véritable « bras armé » des acteurs structurants.

¹⁵⁷ Proposition de Règlement établissant des règles harmonisées concernant l'intelligence artificielle dit Règlement IA, COM(2021) final, 21 avril 2021.

¹⁵⁸ Commission Européen, *Une Europe adaptée à l'ère du numérique. La Commission propose de nouvelles règles et actions en faveur de l'excellence et de la confiance dans l'intelligence artificielle*, Communiqué de presse, 21 avr. 2021.

¹⁵⁹ Sur cette confusion de statut, v. CNPEN, Avis n°3, *Agents conversationnels : enjeux d'éthiques*, novembre 2021, pp. 7-8.

¹⁶⁰ *Business and Professions Code*, Division 7, Part 3, Chapter 6, Section 17940.

¹⁶¹ Pour une analyse critique de cet encadrement a minima, v. *supra*.

II - ASSISTANTS VOCAUX ET ACCÈS AU MARCHÉ

En 2020, la Commission lance une enquête sectorielle sur l'internet des objets grand public au sein de l'Union européenne. Les raisons de cette enquête sont notamment liées à l'expansion rapide de ce secteur ces dernières années et la projection envisagée pour les assistants vocaux dont le nombre devrait doubler d'ici 2024. Le constat y est établi que les consommateurs se méfient des risques induits par l'utilisation de ces technologies sur le respect de leur vie privée, la protection des données personnelles, l'interopérabilité et le manque de compatibilité avec d'autres appareils ou systèmes¹⁶². Le rapport établit également que les acteurs structurants du marché laissent peu de place à l'innovation portée par des initiatives alternatives qui peinent à pénétrer un marché techniquement coûteux mais attractif du fait de l'interaction humain et machine, par la voix. La Commission a donc souhaité comprendre l'environnement des objets connectés et notamment celui de l'assistant vocal, le paysage concurrentiel et les tendances de ce secteur en développement¹⁶³. En réalité, l'effet réseau est particulièrement développé sur ce marché en ce qu'il influence considérablement les négociations entre les différents acteurs d'un marché déployé en écosystème.

L'effet de réseau. Les enjeux concurrentiels sont élevés car les effets de réseau sont indéniables (réputation de la marque, confidentialité, la qualité des applications proposées et le nombre d'utilisateurs) et particulièrement développés à l'échelle de l'assistant vocal. C'est d'ailleurs ce que souligne la House of Representatives considérant que *« les acteurs du marché suggèrent qu'il existe plusieurs barrières à l'entrée pour concurrencer les plateformes générales d'assistants vocaux. Cela inclut de surmonter les effets de réseau dont ont bénéficié les premiers entrants, y compris l'investissement financier dans le matériel, les logiciels et l'infrastructure, et la capacité de vendre des appareils équipés d'assistants vocaux à un prix réduit. Comme de nombreuses plateformes, le marché des assistants vocaux bénéficie d'effets de réseau. Plus il y a d'utilisateurs sur une plateforme, plus il y a de dispositifs et d'applications tiers disponibles, ce qui attire davantage d'utilisateurs. Pour les assistants vocaux, ces effets de réseau sont amplifiés par l'apprentissage automatique et l'intelligence artificielle (IA). Les améliorations du traitement du langage naturel (TALN) et de l'IA devraient améliorer la qualité des assistants vocaux et contribuer à une adoption plus large »*¹⁶⁴.

¹⁶² Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022, point 3 &s.

¹⁶³ Idem, point 10.

¹⁶⁴ United States House of Representatives, *Investigation of competition in Digital Markets*, Majority staff reports and recommendations, Part I, juillet 2022, pp.102-103.

Les enjeux contractuels. Les acteurs structurants occupent des positions stratégiques sur le marché en la forme d'oligopole puissant¹⁶⁵ mais surtout, en capacité de faire adhérer les entreprises utilisatrices à leur environnement, sans exiger d'elles l'exclusivité, sous peine de les ramener au rang des invisibles. Les acteurs structurants disposent d'une capacité de filtrer les contenus, ce qui tient à leur nature de moteur de résultat de l'assistant vocal. Cela contribue à renforcer considérablement le pouvoir de négociation des distributeurs de services (comme les fabricants d'enceintes connectées et autres objets connectés permettant d'étendre l'écosystème ou les fournisseurs d'assistants vocaux). Ces derniers jouent un rôle majeur dans le fonctionnement de l'écosystème car « *ils sont en mesure de capter une part importante de la valeur et d'imposer les conditions de restitution de cette valeur* »¹⁶⁶. De même, les fabricants de solutions techniques peuvent détenir un pouvoir de négociation significatif notamment à l'égard des éditeurs de contenus, sauf les plus attractifs qui pèsent dans le choix d'équipement de l'utilisateur. Dans la majorité des cas, les autres sont étranglés dans la négociation. Le pouvoir de contrainte est fort dès lors que l'éditeur de contenus veut être accessible sur l'un des assistants vocaux largement adoptés par le public afin de toucher une plus grande audience. On leur impose le coût de développement des applications vocales - *skills*, actions, compétences -, le coût de l'accès au catalogue et l'effort technique consenti par les start up et autres PME pour intégrer le système... tout en faisant la promesse d'une collaboration mutuelle et efficace. En réalité, l'asymétrie informationnelle est caractérisée en raison d'une asymétrie de pouvoir. La Commission, dans le cadre de son enquête sectorielle sur les objets connectés grand public, attire l'attention sur le fait « *que les accords de certaines entreprises contiennent des clauses qui amplifient les déséquilibres commerciaux entre elles et la partie contractante la plus faible* »¹⁶⁷. Ce déséquilibre rejoint des problématiques connues s'agissant des magasins d'applications qui ont donné lieu à plusieurs actions sur le fondement du droit de la concurrence¹⁶⁸ ainsi que diverses condamnations au titre du droit des pratiques restrictives de concurrence, notamment en France à l'issue d'actions du Ministre de l'Economie¹⁶⁹. Sur le fondement de la prohibition du déséquilibre significatif posée à l'article L. 442-6, I, 2° du Code de commerce, le Tribunal de commerce de Paris a ainsi jugé illicites plusieurs clauses figurant dans les contrats conclus entre Google LLC, Google Ireland et Google France et des développeurs d'applications, dont celles relatives au prix, les clauses exonératoires de responsabilité ou encore celles

¹⁶⁵ Sur cette qualification s'agissant des assistants virtuels, v. V. Noskova, « Virtual assistants as gatekeepers for consumption? – how information intermediaries shape competition », *European Competition Journal* 2022, DOI: 10.1080/17441056.2022.2129771.

¹⁶⁶ HADOPI / CSA, *Assistants vocaux et enceintes connectées*, 28 mai 2019, pp. 65-66.

¹⁶⁷ Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022, point 489.

¹⁶⁸ V. notamment la procédure pendante devant la Commission européenne à l'encontre d'Apple concernant les règles de l'App Store applicables aux fournisseurs de musique en continu. La Commission interroge deux points en particulier : d'une part, l'obligation d'utiliser le propre mécanisme d'achat intégré d'Apple, imposée aux développeurs pour distribuer leur application de diffusion de musique en continu via l'App Store ; d'autre part, le fait qu'Apple applique aux développeurs certaines restrictions les empêchant d'informer les utilisateurs d'iPhones et d'iPads d'autres possibilités d'achat, moins coûteuses.

¹⁶⁹ Sur l'application cumulée de ce texte avec le DMA une fois celui-ci entré en vigueur, v. M. Behar-Touchais, « L'articulation du DMA avec les autres textes existants », *Concurrence* n°3-2022, p. 51.

permettant une modification du contrat et sa suspension unilatérale ainsi que celles produisant des conditions asymétriques de résiliation¹⁷⁰.

Enfin, le cumul d'intermédiaires puissants entre les éditeurs de contenus et les utilisateurs soulève des questions de captation et de répartition de la valeur à l'égard des éditeurs qui se trouvent en bout de chaîne contractuelle. Dans le cadre de l'étude HADOPI-CSA dédiée aux assistants vocaux et enceintes connectées, la création d'un agrégateur français ou européen est envisagée pour « *rééquilibrer les rapports de force entre les acteurs locaux et les acteurs internationaux* »¹⁷¹. La mise à disposition des contenus culturels en dépend. En effet, il n'existe de coopération qu'entre les acteurs structurants du marché. Parfois la relation contractuelle donne lieu à d'après négociations, parfois le combat est gagné d'emblée par le fournisseur d'assistant vocal dès lors qu'il a verrouillé son système d'exploitation pour ne fournir à l'utilisateur que ses propres produits ou services et des applications pré-installées, ce qui est essentiellement le cas de Siri. Dès lors que les assistants vocaux mettent à disposition des utilisateurs des contenus culturels, on s'interroge sur l'importance d'une coopération entre les fournisseurs de technologie, les éditeurs et ayants droit, de surcroît dans un cadre mondialisé. Ainsi, les négociations contractuelles sont mondiales ce qui multiplie, d'une part, les acteurs économiques plus ou moins présents sur le marché et, d'autre part, le nombre de négociations contractuelles à mener. La Commission constate à cet effet que « *de nombreux fournisseurs de services IoT grand public semblent incapables de négocier des conditions de découvrabilité avec les fournisseurs de plateformes technologiques IoT grand public, qui ne font des exceptions à leurs conditions contractuelles générales que pour les grandes contreparties ayant un pouvoir de négociation important. Certaines parties prenantes craignent que les fournisseurs d'assistants vocaux acceptent - à l'avenir - de ne transporter que les services IoT grand public en échange d'un paiement, ce qui affecterait la rentabilité des fournisseurs de services IoT grand public* »¹⁷².

« **Colonisation** » des écosystèmes. Les liens d'interdépendance existant entre l'assistant vocal, les services essentiels de plateforme et les entreprises utilisatrices, au sens du DMA, lui permettent de « *coloniser les écosystèmes* »¹⁷³, à commencer par le sien avant d'impacter celui des autres acteurs structurants du marché. A ce titre, l'assistant vocal est une pièce majeure de l'écosystème incarnant une nouvelle fonctionnalité qui permet à l'utilisateur de le pénétrer facilement. Cette extension du marché mobilise la Commission européenne qui a d'ailleurs lancé plusieurs pistes pour clarifier la définition du marché et l'adapter au numérique. Elle a publié le 8 novembre 2022 un projet de communication révisé de la Commission destiné à mettre

¹⁷⁰ T. com Paris, 28 mars 2022, n°2018017655 : J.-C. Roda, CCE 2022, étude n°12, n°7 et R. Amaro, *Concurrences* n° 2-2022, Art. n° 106239- v. déjà T. com. Paris, 2 septembre 2019, Amazon, n°2017050625.

¹⁷¹ HADOPI / CSA, Etude préc., p.66.

¹⁷² Commission européenne, Enquête sectorielle préc., point 482.

¹⁷³ Nous reprenons ici l'expression de J.-R. Roda, « Dépasser les limites de l'antitrust », *D.* 2023 (à paraître) - v. également M. E. Stucke, *Breaking Away. How to Regain control over our Data, Privacy and Autonomy*, Oxford University press, 2022.

à jour un texte de vingt-cinq ans¹⁷⁴ qui permet de délimiter le périmètre à l'intérieur duquel la concurrence d'une entreprise s'opère. La Commission propose de nouvelles orientations concernant la définition du marché pour les marchés numériques, par exemple les marchés multifaces et les « écosystèmes numériques » à l'égard notamment des produits construits autour d'un système d'exploitation mobile. Elle donne également davantage d'orientations sur la définition du marché géographique pour un texte prévu pour la fin de l'année 2023. Cependant, ce n'est pas tant la définition du marché qui est impactée que les conséquences juridiques rendant nécessaire l'approche par la régulation *ex ante* en complément du droit de la concurrence principalement *ex post*. En effet, l'assistant vocal apparaît comme une nouvelle fonction sur le marché, une nouvelle forme d'intermédiation, offrant la capacité à l'opérateur d'imposer son pouvoir de négociation et de promouvoir ses propres services concurrents en restreignant l'accès de l'utilisateur aux fonctionnalités de l'appareil par une interface seulement vocale. L'utilisateur n'est donc pas préservé. Il a accès prioritairement à des services et produits mis en avant par la marque qui permet d'identifier fortement l'assistant vocal. Il est peu voire mal informé, soumis à des interfaces trompeuses, des applications préinstallées par défaut, une absence de visibilité de certains contenus, ce qui peut nuire à son accès à la culture et l'enfermer dans ses choix. La réponse unique emporte une véritable restriction de l'exposition des œuvres et des services, voire de l'accès au marché.

Pour autant, ces objets ont vocation à prendre une place dans le quotidien des utilisateurs et tout particulièrement dans leur intimité, à se transformer en bras séculier et agissant de ces derniers, autrement dit en compagnon, majordome, voire en véritable clone prenant les traits de l'avatar. Mais, dans le même temps, l'assistant vocal agit comme le bras armé de l'écosystème. Les assistants vocaux sont à l'écoute de l'utilisateur au sens propre et figuré du terme, opérant au sein d'espaces fermés et contraints comme le domicile et le véhicule, en capacité de capter la voix et les émotions en toutes circonstances pour y répondre de la manière la plus ajustée. C'est l'effet recherché, comme le note l'enquête de la House of Representatives : « *la technologie des assistants vocaux s'améliore plus rapidement lorsque le nombre d'utilisateurs fournissant les échantillons de voix nécessaires à l'entraînement de l'IA est plus élevé. Dans leur témoignage devant le Sous-comité, les professeurs Maurice Stucke et Ariel Ezrachi décrivent ce phénomène comme un "apprentissage par la pratique". Selon ces auteurs, l'effet de réseau de l'apprentissage par la pratique ne se limite pas aux recherches en ligne, mais sera présent dans tout environnement dans lequel les algorithmes évoluent et s'adaptent en fonction de l'expérience, comme par exemple le développement de la reconnaissance vocale ou d'autres instances basées sur l'apprentissage automatique* »¹⁷⁵.

¹⁷⁴ Commission européenne, *Communication sur la définition du marché en cause aux fins du droit communautaire de la concurrence* (97/C 372/03), 9 décembre 1997.

¹⁷⁵ House of Representatives, *Enquête préc.*, pp. 103-104.

L'accès au marché : les enjeux à saisir. Cet écosystème extrêmement intrusif dans sa finalité doit par conséquent être saisi très en amont lors de l'accès au marché afin qu'il soit en capacité de fonctionner convenablement. C'est dans cette perspective qu'aux Etats-Unis l'American Innovation and Choice Online Act vise à encourager l'interopérabilité et interdire l'autopréférence discriminatoire et inéquitable. Le législateur européen est allé plus loin dans la démarche, reconnaissant que l'assistant vocal ne doit pas se positionner comme un outil de « colonisation » des écosystèmes pour étouffer le pouvoir de négociation de ses partenaires, bloquer le consommateur dans son écosystème ou le « nudger » pour l'amener dans son écosystème. Alors que le droit des pratiques anticoncurrentielles et restrictives de concurrence demeure applicable¹⁷⁶ mais ne permet pas d'éviter ces comportements, il convient donc d'imposer aux assistants vocaux des obligations *a priori* pour assurer l'équité et la contestabilité des marchés numériques. Le DMA s'est saisi de ces enjeux par son approche transversale. Il prévoit à cet effet des familles d'obligations pesant sur les contrôleurs d'accès, à partir desquelles on distingue les obligations générales posées à l'article 5, complétées par une série d'obligations spécifiques que l'on retrouve éparpillées dans le texte.

Par conséquent, les contrôleurs d'accès sont notamment invités à ne plus imposer par défaut l'installation de leur système d'exploitation. Un écran multi-choix devra être proposé aux utilisateurs pour être en capacité d'opter pour un service concurrent (article 6.3). Ils ne pourront plus non plus mettre en avant leurs services et produits par rapport à ceux des vendeurs qui utilisent leur plateforme (article 6.5) ou exploiter les données des vendeurs pour les concurrencer (article 6.2) ni imposer aux développeurs d'applications certains services annexes comme le système de paiement par exemple (article 6.4). Si les assistants virtuels sont qualifiés de services de plateforme essentiels au sens du DMA, des obligations pourront s'imposer à eux telle que l'obligation d'interopérabilité gratuite (article 6.7), l'interdiction de combiner les données personnelles des utilisateurs provenant d'autres services (article 5.2), l'obligation de communiquer leurs chiffres aux annonceurs et aux éditeurs (article 5.9 et 10), ou bien encore l'obligation de permettre la portabilité dynamique et en temps réel des données (article 6.9). L'article 14 prévoit en outre une obligation d'informer sur les concentrations. Ces services pourront également être soumis à une enquête de marché en vertu du chapitre IV. Précisons que les obligations posées à l'article 5 du DMA sont transversales, en ce qu'elles visent l'ensemble des entreprises désignées comme contrôleur d'accès là où la liste des obligations figurant à l'article 6 permettra à la Commission de les mettre au cas par cas à la charge de ces opérateurs.

¹⁷⁶ Sur cette question, v. M. Behar-Touchais, « L'articulation du DMA avec les autres textes existants », *préc.*

Règlement 2022/1925, article 5 : Obligations incombant aux contrôleurs d'accès

1. Le contrôleur d'accès se conforme à toutes les obligations énoncées au présent article pour chacun de ses services de plateforme essentiels énumérés dans la décision de désignation conformément à l'article 3, paragraphe 9.

2. Tout contrôleur d'accès est tenu de ne pas :

a) traiter, aux fins de la fourniture de services de publicité en ligne, les données à caractère personnel des utilisateurs finaux qui recourent à des services de tiers utilisant des services de plateforme essentiels fournis par le contrôleur d'accès ;

b) combiner les données à caractère personnel provenant du service de plateforme essentiel concerné avec les données à caractère personnel provenant de tout autre service de plateforme essentiel ou de tout autre service fourni par le contrôleur d'accès, ni avec des données à caractère personnel provenant de services tiers ;

c) utiliser de manière croisée les données à caractère personnel provenant du service de plateforme essentiel concerné dans le cadre d'autres services fournis séparément par le contrôleur d'accès, y compris d'autres services de plateforme essentiels, et inversement ; et

d) inscrire les utilisateurs finaux à d'autres services du contrôleur d'accès dans le but de combiner des données à caractère personnel, à moins que ce choix précis ait été présenté à l'utilisateur final et que ce dernier ait donné son consentement au sens de l'article 4, point 11), et de l'article 7 du règlement (UE) 2016/679.

Lorsque le consentement donné aux fins du premier alinéa a été refusé ou retiré par l'utilisateur final, le contrôleur d'accès ne réitère pas sa demande de consentement pour la même finalité plus d'une fois par période d'un an.

Le présent paragraphe est sans préjudice de la possibilité pour le contrôleur d'accès de se fonder sur l'article 6, paragraphe 1, points c), d) et e), du règlement (UE) 2016/679, le cas échéant.

3. Le contrôleur d'accès n'empêche pas les entreprises utilisatrices de proposer les mêmes produits ou services aux utilisateurs finaux au moyen de services d'intermédiation en ligne tiers ou de leur propre canal de vente directe en ligne à des prix ou conditions différents de ceux qui sont proposés par les services d'intermédiation en ligne du contrôleur d'accès.

4. Le contrôleur d'accès permet aux entreprises utilisatrices de communiquer et de promouvoir leurs offres gratuitement, y compris à des conditions différentes, auprès des utilisateurs finaux acquis grâce à son service de plateforme essentiel ou via d'autres canaux, et de conclure des contrats avec ces utilisateurs finaux, en utilisant ou non à cette fin les services de plateforme essentiels du contrôleur d'accès.

5. Le contrôleur d'accès permet aux utilisateurs finaux, par l'intermédiaire de ses services de plateforme essentiels, d'accéder à des contenus, abonnements, fonctionnalités ou autres éléments et de les utiliser en se servant de l'application logicielle de l'entreprise utilisatrice, y compris lorsque ces utilisateurs finaux ont acquis de tels éléments auprès des entreprises utilisatrices concernées sans avoir recours aux services de plateforme essentiels du contrôleur d'accès.

6. Le contrôleur d'accès n'empêche ni ne restreint directement ou indirectement la possibilité pour les entreprises utilisatrices ou les utilisateurs finaux de faire part à toute autorité publique compétente, y compris les juridictions nationales, de tout problème de non-respect, par le contrôleur d'accès, du droit de l'Union ou national pertinent dans le cadre des pratiques de ce dernier. Cela s'entend sans préjudice du droit des entreprises utilisatrices et des contrôleurs d'accès d'établir, dans leurs accords, les conditions d'utilisation de mécanismes légaux de traitement des plaintes.

7. Le contrôleur d'accès n'exige pas des utilisateurs finaux qu'ils utilisent, ni des entreprises utilisatrices qu'elles utilisent, proposent ou interagissent avec un service d'identification, un navigateur internet ou un service de paiement, ou un service technique qui appuie la fourniture des services de paiement, tels que des systèmes de paiement destinés aux achats dans des applications, de ce contrôleur d'accès dans le cadre des services fournis par les entreprises utilisatrices en ayant recours aux services de plateforme essentiels de ce contrôleur d'accès.

8. Le contrôleur d'accès n'exige pas des entreprises utilisatrices ou des utilisateurs finaux qu'ils s'abonnent ou s'enregistrent à tout autre service de plateforme essentiel énuméré dans la décision de désignation conformément à l'article 3, paragraphe 9, ou atteignant les seuils visés à l'article 3, paragraphe 2, point b), comme condition pour être en mesure d'utiliser l'un des services de plateforme essentiels de ce contrôleur d'accès énumérés en vertu dudit article, d'y accéder, de s'y inscrire ou de s'y enregistrer.

9. Le contrôleur d'accès communique quotidiennement à chaque annonceur à qui il fournit des services de publicité en ligne, ou aux tiers autorisés par les annonceurs, à la demande de l'annonceur, des informations gratuites relatives à chaque publicité mise en ligne par l'annonceur, en ce qui concerne :

a) le prix et les frais payés par cet annonceur, y compris les déductions et suppléments éventuels, pour chacun des services de publicité en ligne concernés fournis par le contrôleur d'accès ;

b) la rémunération perçue par l'éditeur, y compris les déductions et suppléments éventuels, sous réserve du consentement de l'éditeur ;

c) et les mesures quantitatives à partir desquelles chacun des prix, frais et rémunérations est calculé.

Dans le cas où un éditeur ne consent pas au partage d'informations sur la rémunération perçue, comme visé au point b) du premier alinéa, le contrôleur d'accès fournit gratuitement à chaque annonceur des informations sur la rémunération moyenne quotidienne perçue par

cet éditeur, y compris les déductions et suppléments éventuels, pour les publicités concernées.

10. Le contrôleur d'accès communique quotidiennement à chaque éditeur à qui il fournit des services de publicité en ligne, ou aux tiers autorisés par les éditeurs, à la demande de l'éditeur, des informations gratuites relatives à chaque publicité affichée dans l'inventaire de l'éditeur, en ce qui concerne :

- a) la rémunération perçue et les frais payés par cet éditeur, y compris les déductions et suppléments éventuels, pour chacun des services de publicité en ligne concernés fournis par le contrôleur d'accès ;
- b) le prix payé par l'annonceur, y compris les déductions et suppléments éventuels, sous réserve du consentement de l'annonceur ;
- c) et la mesure à partir de laquelle chacun des prix, frais et rémunérations est calculé.

Dans le cas où un annonceur ne consent pas au partage d'informations, le contrôleur d'accès fournit gratuitement à chaque éditeur des informations sur le prix moyen quotidien payé par cet annonceur, y compris les déductions et suppléments éventuels, pour les publicités concernées.

Afin de saisir l'ensemble des enjeux relatifs au marché des assistants vocaux, doivent être examinées plus spécifiquement les points névralgiques qui permettront un accès non discriminatoire et équitable au marché de manière à garantir la liberté de choix de l'utilisateur, notamment lors de son accès aux contenus culturels. Il convient ainsi d'interroger l'autopréférence sous toutes ses facettes (A), l'interopérabilité des systèmes et des applications (B) ainsi que l'accès aux données techniques et celles des utilisateurs (C).

A - L'autopréférence (ne pas bloquer l'accès au marché)

L'autopréférence : un enjeu majeur. S'agissant des fournisseurs d'assistants vocaux, l'une des principales préoccupations tient aux possibilités d'autopréférence qui peuvent, comme cela a été évoqué *supra*, résulter de deux dimensions. Tout d'abord, l'enceinte de l'opérateur d'assistant vocal est utilisée de préférence pour interagir avec l'assistant vocal. Ensuite, les applications privilégiées pour répondre aux requêtes de l'utilisateur seront celles du même opérateur. Enfin, l'absence d'écran dans l'interaction avec les assistants vocaux ainsi que les défis liés à la réponse unique sur ces systèmes peuvent conduire à favoriser les services de ces opérateurs par défaut.

Le rapport final de l'enquête de la Commission européenne sur l'loD fait part des nombreuses préoccupations des parties prenantes à cet égard. Celles-ci portent notamment sur la pré-installation de l'assistant vocal et des applications, les paramètres par défaut définis par les fournisseurs et la mise en évidence de produits et services. Ainsi, s'agissant des fonctionnalités « prêtes à l'emploi » offertes aux utilisateurs, les répondants indiquent que de telles pratiques « *ont un impact sur la possibilité de découvrir les services IoT grand public, au point de constituer un (dés)avantage concurrentiel* »¹⁷⁷ et que les grands fournisseurs de services de contenu créatif, actifs à l'échelle mondiale, profiteraient particulièrement de cette situation. A ce titre, il est souligné que les pratiques de pré-installation sont très répandues - l'assistant vocal est généralement programmé pour utiliser par défaut un service spécifique pour réaliser une activité spécifique - qui conduisent le plus souvent les fournisseurs d'assistants vocaux à diriger l'utilisateur vers leurs propres services connexes (par exemple musique, podcasts, livres, contenus audiovisuels, achats). Certains répondants expliquent alors que « *la pré-installation constitue un désavantage concurrentiel pour les fournisseurs de services IoT grand public dont les applications ne sont pas préinstallées, car un utilisateur doit effectuer des démarches supplémentaires pour accéder à un service qui n'est pas inclus dans l'offre "out of the box". À l'inverse, les services préinstallés bénéficient de niveaux plus élevés de découvrabilité par les utilisateurs* »¹⁷⁸ ; ils relèvent également que « *les services par défaut sont plus 'collants', c'est-à-dire qu'ils incitent les utilisateurs à rester plus longtemps sur le service par défaut* »¹⁷⁹ et que *l'accès à un autre fournisseur de service supposera d'ajouter un nom d'invocation spécifique à sa commande vocale* » (par exemple « *joue le morceau X sur le service Y* »). Sont également envisagés les effets des pratiques de prééminence déterminant la visibilité et la trouvabilité d'un service. Ces pratiques sont soumises à d'importantes négociations¹⁸⁰.

Le rapport de la Commission souligne encore l'existence de pratiques d'exclusivité et de vente liée pour les assistants vocaux, ainsi que de pratiques limitant la possibilité d'utiliser différents assistants vocaux sur le même dispositif intelligent¹⁸¹. Il cite en particulier le cas de l'exclusivité de la présence d'un assistant vocal sur un appareil intelligent. Les résultats de l'enquête révèlent à cet égard que les fabricants de ces objets proposent en règle générale aux utilisateurs des appareils avec un seul assistant vocal intégré alors qu'il existe une demande de la part des utilisateurs pour une double commande de l'appareil par un assistant vocal généraliste (modalité la plus courante) et un assistant vocal spécialisé (ce qui est soulevé par la concurrence). La Commission précise à cet effet que « *l'utilisation simultanée des assistants vocaux, à savoir le passage d'un assistant vocal à l'autre en utilisant un mot d'activation spécifique pour activer l'un des assistants vocaux, n'est possible que sur un nombre*

¹⁷⁷ Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022, point 440.

¹⁷⁸ Commission européenne, *Rapport préc.*, point 450.

¹⁷⁹ Commission européenne, *Rapport préc.*, point 458.

¹⁸⁰ Commission européenne, *Rapport préc.*, points 464 & s.

¹⁸¹ Commission européenne, *Rapport préc.*, points 470 & s.

limité d'appareils intelligents fabriqués par les répondants. Certaines préoccupations concernant l'impossibilité de permettre l'utilisation simultanée des assistants vocaux sur les appareils intelligents ont été soulevées lors de l'enquête sectorielle ». En l'état, c'est l'une des préoccupations des acteurs qui se positionnent sur le marché comme une simple interface vocale. Or, le mot d'activation spécifique est lié à l'identification de la marque parfois au détriment de la marque des objets connectés déclenchés par la commande de l'assistant vocal. Par conséquent, ces pratiques peuvent avoir pour effet de renforcer l'identité de l'assistant vocal dans l'environnement domestique. Les services audiovisuels ont des préoccupations similaires, voire supplémentaires. Outre l'enfermement de l'utilisateur dans un écosystème, cela pose à terme la question de l'avenir de la télévision et de la radio hertzienne gratuite et universelle, selon l'évolution des interfaces vocales et des technologies utilisées si l'assistant vocal était amené à devenir le centre de commande des services audiovisuels.

Le rapport relève que des craintes résultent également du risque de désintermédiation¹⁸². Un tel risque tient à la position des assistants vocaux et des systèmes d'exploitation de dispositifs intelligents en tant qu'intermédiaire entre les utilisateurs et les dispositifs intelligents ou les services de l'IdO pour les consommateurs dans la mesure où cette position, associée à leur rôle clé dans la production et la collecte de données, leur permettrait de contrôler les relations avec les utilisateurs. Dans ce contexte, les parties prenantes ont également fait part de leurs préoccupations quant à la découvrabilité et à la visibilité de leurs services en matière d'IdO pour les consommateurs. Ils craignent alors une perte de la reconnaissance de leur marque et de leur relation directe avec les utilisateurs, ainsi que « *l'absence d'accès cohérent et immédiat aux données pertinentes sur l'utilisation de leurs services* », ce qui les empêche de personnaliser l'expérience client en temps réel et les place dans « *une situation de désavantage concurrentiel par rapport aux services connexes offerts par le fournisseur de l'assistant vocal* »¹⁸³.

Pratique(s) d'autopréférence. Différentes modalités d'autopréférence peuvent ainsi être identifiées. *Stricto sensu*, l'autopréférence va renvoyer à l'autoréférence - ou autopréférence dans le classement - qui vise le cas où l'assistant vocal donnerait la priorité aux produits ou services connexes offerts par l'entreprise lui fournissant l'assistant vocal. *Largo sensu*, l'autopréférence peut permettre de saisir les cas de vente groupée, à savoir lorsqu'un ensemble de biens et services sont fournis en un seul lot. Dans le cas des assistants vocaux, cette pratique pourra consister à préinstaller un assistant sur le matériel *hardware* (smartphone, ordinateur, enceinte) de l'utilisateur, ou encore à préinstaller diverses applications vocales sur l'assistant vocal lors de sa fourniture à l'utilisateur. Cette pratique peut être renforcée par le paramétrage des valeurs par défaut consistant par exemple, lors de la requête utilisateur, à privilégier le lancement de l'application située dans l'écosystème propriétaire du fournisseur d'assistant vocal. L'autopréférence peut encore résulter de

¹⁸² Commission européenne, Rapport préc., points 474 & s.

¹⁸³ Commission européenne, Rapport final préc., point 483.

pratiques d'exclusion de concurrents ou de pratiques les défavorisant au profit des services connexes de l'entreprise fournissant l'assistant vocal. Ces pratiques peuvent notamment être de nature tarifaire ou relever de traitements des données dans la mesure où l'opérateur de plateforme s'appuie sur les données de ses utilisateurs professionnels afin d'améliorer ses propres offres pour les concurrencer en qualité de fournisseur de service ou vendeur tiers.

L'autopréférence sanctionnée par le droit de la concurrence. Plusieurs actions ont été menées sur le fondement du droit de la concurrence visant spécifiquement les différentes pratiques d'autopréférence des grands opérateurs du numérique. Outre le contentieux relatif à Microsoft¹⁸⁴, il convient de mentionner à cet égard la condamnation de Google dans le cadre de l'affaire Google Shopping pour avoir manipulé artificiellement l'algorithme de son moteur de recherche afin de privilégier son propre comparateur, ce qui confirme que l'autopréférence peut, en fonction des circonstances, constituer un abus de position dominante¹⁸⁵ ; également dans l'affaire Google Android, s'agissant cette fois de la pré-installation des navigateur et moteur de recherche imposée contractuellement aux fournisseurs d'appareils mobiles¹⁸⁶. On mentionnera en outre l'ouverture d'une enquête s'agissant de l'obligation imposée aux fournisseurs de matériels fonctionnant sous Android de préinstaller Google Assistant à l'exclusion du service fourni par ses concurrents¹⁸⁷. La Commission européenne a également « *ouvert de nouveaux fronts* »¹⁸⁸ en visant particulièrement les pratiques d'Amazon et celles d'Apple. Concernant Amazon, une première communication des griefs a été envoyée par la Commission s'agissant de l'utilisation des données non publiques des utilisateurs professionnels de sa place de marché afin de favoriser ses propres produits¹⁸⁹. A cela, s'est ajouté le lancement d'une enquête relative à la « buy box » et d'Amazon Prime afin d'analyser la pratique d'autopréférence consistant à privilégier sa propre activité de vente au détail ou celles des distributeurs utilisant les services logistiques de la place de marché¹⁹⁰. En juillet 2022 - soit à la veille de l'adoption du DMA -, Amazon a fait une proposition d'engagement à la Commission européenne qui répondait, entre autres, à la préoccupation de la Commission concernant l'auto-référencement lors de l'attribution de la case d'achat¹⁹¹. En outre, la Commission a adressé une communication des griefs à Apple lui reprochant d'obliger les fournisseurs d'applications de recourir au mécanisme d'achat intégré d'Apple afin de distribuer leur application de streaming musical. De plus, la Commission a relevé que l'opérateur limitait la capacité des fournisseurs d'applications à utiliser des fournisseurs de paiement alternatif in-app et les obligeait à utiliser son propre système (Apple Pay) tout en leur imposant de verser des commissions auxquelles n'était pas

¹⁸⁴ TPI, 17 septembre 2007, Microsoft Corp. c. Commission, aff. T-201/04.

¹⁸⁵ Trib. UE, 10 novembre 2021, Google Alphabet c. Commission, aff. T-612/17.

¹⁸⁶ Trib. UE, 14 septembre 2022, Google Alphabet c. Commission, aff. T-604/18.

¹⁸⁷ F. Y. Chee and N. Balu, « Google's Voice assistant in new EU antitrust investigation », *Reuters*, 9 septembre 2021.

¹⁸⁸ J.-C. Roda, « Un an de droit de la concurrence dans l'univers numérique », *CCE 2022*, étude n°12, n°5 qui décrit ensuite ces différents contentieux.

¹⁸⁹ Affaire COMP/AT.40462.

¹⁹⁰ Affaire COMP/AT.40703.

¹⁹¹ https://ec.europa.eu/competition/antitrust/cases1/202229/AT_40462_8414012_7971_3.pdf

soumis son propre service d'abonnement musical Apple Play. Cela emportait donc un déficit d'attractivité des applications tierces¹⁹².

L'autopréférence encadrée par le DMA : do's and don'ts. Dans le prolongement de ces différents contentieux et afin de garantir la contestabilité et l'équité des marchés numériques, le DMA vient saisir ces pratiques d'autopréférence selon diverses modalités, suivant la méthode du « *do's* » *and don'ts* » caractérisant le texte¹⁹³. Il s'intéresse tout particulièrement aux mécanismes de classement des produits et services mis à la disposition de l'utilisateur dans le cadre de la réponse unique, mais également aux pratiques visant à installer des fonctionnalités et des applications par défaut.

A titre préventif, le texte interdit tout d'abord aux contrôleurs d'accès de pratiquer l'autoréférencement, au sens de l'autopréférence dans le classement, en visant expressément les assistants virtuels. Ainsi, son article 6.5 prévoit que « *Le contrôleur d'accès n'accorde pas, en matière de classement ainsi que pour l'indexation et l'exploration qui y sont liées, un traitement plus favorable aux services et produits proposés par le contrôleur d'accès lui-même qu'aux services ou produits similaires d'un tiers. Le contrôleur d'accès applique des conditions transparentes, équitables et non discriminatoires à ce classement* »¹⁹⁴. Par ailleurs, la règle dite d'anti-contournement de l'article 13 du DMA pourra conduire à prohiber l'autoréférencement et autres distorsions de recommandations tendant à imposer un certain comportement aux utilisateurs du service du contrôleur d'accès en ce que cette disposition prévoit que « *Le contrôleur d'accès ne détériore ni les conditions, ni la qualité de l'un de ses services de plateforme essentiels fournis aux entreprises utilisatrices ou aux utilisateurs finaux qui font valoir leurs droits ou choix prévus aux articles 5, 6 et 7, et ne rend pas l'exercice de ces droits ou choix excessivement difficile, y compris en proposant des choix à l'utilisateur final de manière partielle, ou encore en utilisant la structure, la conception, la fonction ou le mode de fonctionnement d'une interface utilisateur ou d'une partie connexe pour perturber l'autonomie des utilisateurs finaux ou des entreprises utilisatrices, leur prise de décision ou leur libre choix* »¹⁹⁵. Au cours des négociations du texte, certains ont pu préconiser que l'interdiction d'autopréférence « *soit étendue au-delà du classement des offres à toute technique permettant d'influencer les utilisateurs finaux pour les orienter vers les produits et services du contrôleur d'accès et d'entreprises liées* »¹⁹⁶. Tel paraît être le cas *a minima* lorsque le DMA, en son article 6.2, limite l'utilisation des données par le

¹⁹² Aff. COMP/AT.40437.

¹⁹³ L. Idot, « Propos introductifs », *Concurrence* n°3-2020, p. 45, n°20.

¹⁹⁴ Le classement est défini par l'article 2.22 du DMA comme « *la priorité relative accordée aux biens ou services proposés par le biais de services d'intermédiation en ligne, de services de réseaux sociaux en ligne, de services de plateformes de partage de vidéos ou d'assistants virtuels, ou la pertinence reconnue aux résultats de recherche par les moteurs de recherche en ligne, tels qu'ils sont présentés, organisés ou communiqués par les entreprises fournissant des services d'intermédiation en ligne, des services de plateformes de partage de vidéos, des assistants virtuels ou des moteurs de recherche en ligne, indépendamment des moyens technologiques utilisés pour une telle présentation, organisation ou communication et indépendamment du fait qu'un seul résultat soit ou non présenté ou communiqué* ».

¹⁹⁵ En ce sens, v. M. Peitz, *The prohibition of self-preferencing in the DMA*, Issue Paper, CERRE, novembre 2022, p. 10.

¹⁹⁶ Sénat, *Résolution européenne sur la proposition de Règlement sur les marchés numériques (DMA)*, n°32, 12 novembre 2021, point 49.

contrôleur d'accès en concurrence avec les entreprises utilisatrices de son service ; il vise alors les données « *quelles qu'elles soient, qui ne sont pas accessibles au public, qui sont générées ou fournies par ces entreprises utilisatrices dans le cadre de leur utilisation des services de plateforme essentiels concernés ou des services fournis conjointement aux services de plateforme essentiels concernés, ou à l'appui de ceux-ci, y compris les données générées ou fournies par les clients de ces entreprises utilisatrices* ». Il en va de même de l'interdiction des interfaces de choix trompeuses ou manipulatrices - ou *dark patterns* - posée à l'article 13, qui s'appliquera au-delà du seul classement (v. *infra*). Au-delà de ces interdictions, le texte se saisit d'un certain nombre d'autres pratiques mais cette fois sous la forme d'obligations positives qui consistent à imposer aux contrôleurs d'accès de limiter le verrouillage de leurs utilisateurs par le mécanisme du paramétrage par défaut s'agissant de la pré-installation et des valeurs par défaut pour leur garantir une liberté de choix (v. *infra*) ; ceci justifie également de leur imposer des mesures d'interopérabilité et de garantir l'accès aux données.

En réalité, les pratiques d'autopréférence constituent le risque majeur dont il convient de se saisir en priorité afin d'assurer la fluidité du marché en posant les bonnes barrières de prévention et de protection. Cette démarche est particulièrement complexe, car elle suppose de se saisir de toutes les pratiques émergentes sur le marché de manière à faire évoluer la démarche préventive du DMA. L'objectif du texte consiste en effet à ouvrir les écosystèmes tout en assurant leur fonctionnement équitable, en identifiant quelques pratiques *a priori* qui contribuent à les verrouiller. Par sa finalité préventive et transversale, l'approche du DMA va en cela s'atteler aux pratiques en capacité de déséquilibrer structurellement l'écosystème. Ces pratiques ne peuvent être définitivement figées dans une liste d'obligations. A cet effet, le DMA prévoit que les obligations posées à l'article 6 évolueront conformément aux mécanismes prévus à l'article 8. Cette démarche préventive s'accompagnera d'un second niveau d'intervention *ex post* pour sanctionner, cette fois-ci sur le fondement du droit de la concurrence, les déséquilibres conjoncturels du marché passés au travers des mailles du premier niveau de la régulation.

B - L'interopérabilité (pour accéder au marché)

Pourquoi l'interopérabilité ? Le coût des investissements technologiques, les importants effets de réseau (principalement indirects), ainsi que les nombreux autres effets de verrouillage à l'œuvre (par ex. l'impossibilité pour les utilisateurs de récupérer leurs données) constituent des obstacles indéniables à l'entrée ou l'expansion sur le marché des assistants vocaux polyvalents à court terme. A cet égard, on relèvera qu'au côté des économies d'échelle et d'envergure (couvertes par la mention aux investissements nécessaires), les effets de réseau (lorsqu'ils sont captés par les acteurs en place) constituent des sources classiques de défaillance de marché en économie industrielle. Une attention particulière doit être portée ici aux effets de

réseau indirects dès lors qu'ils sont particulièrement puissants en ce qui concerne les assistants vocaux. En effet, une part importante d'utilisateurs finals augmente la valeur du service pour les fournisseurs d'applications (et inversement). Cela rend les quelques assistants vocaux sur le marché incontournables pour les deux groupes d'utilisateurs ; d'autant qu'il est compliqué pour les utilisateurs d'utiliser plusieurs assistants vocaux (v. *supra*). Dans cette situation, l'interopérabilité permet de limiter la capacité que peuvent avoir les quelques fournisseurs d'assistants vocaux de contrôler indûment l'accès à leur infrastructure incontournable.

Dans le cadre de son enquête sur les marchés numériques, la House of Representatives illustre le phénomène de la manière suivante :

« À mesure que le marché des assistants vocaux se développe, il peut être difficile pour les utilisateurs de passer d'une plateforme à l'autre. Les plateformes d'assistants vocaux n'étant pas toujours interopérables, les utilisateurs devraient supporter des coûts pour acheter un ou plusieurs nouveaux appareils. En outre, la technologie des assistants vocaux est conçue pour apprendre les préférences de l'utilisateur au fil du temps. Ces préférences vont des paramètres tels que les informations de facturation et les services par défaut pour répondre aux commandes de musique à un apprentissage plus avancé comme les commandes vocales passées et l'historique des achats. Au fur et à mesure qu'un assistant vocal améliore sa " compréhension " de son utilisateur, il peut augmenter les coûts associés au passage à une autre plateforme. Comme l'a noté un acteur du marché dans une soumission au sous-comité, "l'utilisateur peut devenir plus dépendant de cet assistant vocal particulier et être beaucoup moins enclin à utiliser un assistant vocal rival qui n'a pas encore "rattrapé" ses préférences".

La conception de la plupart des assistants vocaux, en particulier sur les appareils sans écran amplifie la capacité des plates-formes d'assistants vocaux à favoriser leurs services par défaut ou en tant que réponse avec un choix limité. Cette dynamique permet aux assistants vocaux les plus populaires de favoriser leurs services de premier plan. Il existe également un potentiel important d'utilisation abusive des données pour nuire à la concurrence ou aux consommateurs. (...)

Enfin, les leaders de l'écosystème des assistants vocaux fixent les règles pour les tiers. Pour fabriquer un appareil doté d'un assistant vocal, les acteurs du marché doivent se conformer aux spécifications des fournisseurs de plateformes d'assistants vocaux.

Comme l'a noté M. Spence de Sonos dans son témoignage devant le sous-comité : Pour avoir accès à leurs plates-formes et s'intégrer à leurs services, ces entreprises émettent toutes sortes de demandes à prendre ou à laisser, allant d'un accès précoce et techniquement détaillé à nos feuilles de route de produits, à des données commerciales exclusives, y compris les prévisions de ventes, à des renoncements à des droits contractuels essentiels. Le sous-comité a également entendu de nombreux développeurs d'assistants vocaux qui ont lutté pour obtenir l'accès aux fonctionnalités clés nécessaires à la création de leurs applications, telles que les commandes non traitées de l'utilisateur. Bien qu'il soit encore en développement, le marché des assistants vocaux montre les premiers signes de concentration du marché »¹⁹⁷.

Source: House of Representatives, Investigation of competition in Digital Markets, majority staff reports and recommandations, Part I, juillet 2022, p. 104 &s

Définition de l'interopérabilité. Le terme même d'interopérabilité couvre des réalités multiples mais de manière générale, « *c'est la capacité que possède un produit du*

¹⁹⁷ House of Representatives, *Investigation of competition in Digital Markets*, majority staff reports and recommandations, Part I, juillet 2022, p. 101&s.

« système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et sans restriction d'accès ou de mise en œuvre »¹⁹⁸, sans effort particulier de l'utilisateur. Dans le cadre d'une approche simplifiée et globale de l'interopérabilité, le PeREN la décline sous trois aspects relevant à la fois de l'interopérabilité proprement dite, couvrant sa forme horizontale ou verticale, de la compatibilité des systèmes et de la portabilité des données. Le tout se retrouvant de manière très graduée au sein des écosystèmes allant d'une absence totale d'interopérabilité, à une interopérabilité complète, en passant par une interopérabilité ne concernant que certaines fonctionnalités partagées horizontalement par tous les acteurs¹⁹⁹.

PeREN : Définition de l'interopérabilité

« L'**interopérabilité horizontale** (entre des produits concurrents, comme par exemple entre des médias sociaux)²⁰⁰ et l'**interopérabilité verticale** (interopérabilité avec des produits complémentaires, par exemple entre une place de marché et des outils tiers utilisés par les vendeurs présents sur la plateforme) ».

« La **compatibilité** est une forme limitée d'interopérabilité, qui permet à deux systèmes spécifiques, mais de types différents, de communiquer entre eux ».

« La **portabilité** des données²⁰¹, quant à elle, concerne uniquement la migration des données entre deux services ou deux plateformes ».

Source : PeREN, *Eclairage sur : l'interopérabilité*, octobre 2021, p.1.

Interopérabilité : quels objectifs ? La performance de l'assistant vocal dépend des applications à partir desquelles il va proposer une réponse à la commande de l'utilisateur. Or, le développement de nouveaux produits et services se heurte à l'intégration verticale des principaux fournisseurs, qui combinent leurs propres produits et services à ceux des tiers pour proposer une offre attractive. Dès lors, la capacité à interconnecter les différentes composantes d'un écosystème dans le cadre de liens interdépendants est particulièrement questionnée. Les différentes composantes de l'écosystème doivent être en capacité de communiquer entre elles, ce qui suppose de garantir l'interopérabilité entre les dispositifs connectés, les

¹⁹⁸ Association francophone des utilisateurs de logiciels libres cité par le PeREN, *Eclairage sur : l'interopérabilité*, octobre 2021, p.1 - v. également Direction Interministérielle du Numérique et du Système d'Information et de Communication de l'Etat, « Référentiel Général d'Interopérabilité : Standardiser, s'aligner et se focaliser pour échanger efficacement », Version 2.0, 2015 - comp. article 2 DMA définissant l'interopérabilité comme « la capacité d'échanger des informations et d'utiliser mutuellement les informations échangées par le biais d'interfaces ou d'autres solutions, de telle sorte que tous les éléments du matériel informatique ou des logiciels fonctionnent de toutes les manières dont elles sont censées fonctionner avec d'autres matériels informatiques et logiciels ainsi qu'avec les utilisateurs ».

¹⁹⁹ PeREN, Rapport préc.

²⁰⁰ L'interopérabilité horizontale est davantage un remède aux problèmes liés aux effets de réseaux directs qui ne semble pas totalement pertinent dans ce contexte (cela ne nécessite pas de modification, l'interopérabilité horizontale n'est pas trop évoquée).

²⁰¹ Au-delà des effets de réseaux, la portabilité dynamique (qui constitue une forme d'interopérabilité) telle que prévue par le DMA (article 6.9) pourrait aussi présenter un intérêt. Ce remède n'est pas lié aux effets de réseau mais plus à d'autres effets de verrouillage, du type l'impossibilité de récupérer ses données (en temps réel) pour rejoindre un autre service.

assistants vocaux et les services proposés aux utilisateurs. Sans interopérabilité, il est impossible de proposer aux utilisateurs des produits et services diversifiés faisant ainsi obstacle à l'autopréférence. Outre les fournisseurs d'applications, il serait aussi intéressant d'ajouter que l'interopérabilité permet aux utilisateurs finals, de l'autre côté de la plateforme, de bénéficier des effets de réseau indirects (c'est-à-dire une plus large gamme de services sur la plateforme, donc plus de choix, et en théorie un effet positif sur la qualité, le prix, voire la sécurité et la confidentialité si cela entre dans leurs critères de choix...). Par ailleurs, l'interopérabilité permet à de nouveaux entrants sur un marché de profiter des avantages de « l'effet réseau » existant par le jeu des acteurs structurants, pour l'utiliser à leur propre avantage et se hisser à leur tour sur le marché²⁰². C'est généralement le fait de la disponibilité d'un protocole ou d'un standard.

En cela, l'interopérabilité est le meilleur levier d'une régulation *ex ante* asymétrique²⁰³, mais ce n'est qu'un outil parmi d'autres dont l'impact devra être évalué de manière contextuelle à la fois dans le cadre de ses enjeux économiques (effet levier, coût d'entrée sur le marché, modèle économique), techniques (difficultés à ouvrir des interfaces, sécurité des données) mais aussi juridiques (aspects concurrentiels, contractuels, portabilité des données, protection des droits de propriété intellectuelle).

Constat de l'absence d'interopérabilité entre assistants vocaux, et entre assistants vocaux et entreprises utilisatrices tierces. C'est le constat réalisé par la Commission européenne laquelle, dans le cadre de son enquête sectorielle sur l'internet des objets grand public²⁰⁴, met au centre de ses développements les questions d'interopérabilité. Il s'agit de l'élément technique conditionnant la création des écosystèmes et le principal levier d'action pour les ouvrir dans la perspective de garantir l'accès au marché. La Commission identifie deux nœuds bloquant l'accès au marché. D'une part, l'opérateur du système contrôle la possibilité d'intégrer le marché par des normes techniques propriétaires, et d'autre part il va contrôler les fonctionnalités des assistants vocaux pour en limiter l'accès.

Les assistants vocaux sont arrivés sur le marché avant que ne soit envisagée la définition d'un standard ou d'un protocole de communication. La prédominance des technologies propriétaires a conduit parfois à la création de normes « de fait », accompagnées de processus de certification. Ces quelques acteurs structurants contrôlent unilatéralement les processus d'interopérabilité et d'intégration et sont en mesure d'en limiter les fonctionnalités. Les spécificités techniques et les spécifications de logiciels pour permettre l'accès aux assistants vocaux sont mis à la disposition de tiers sous réserve de la conclusion d'accords d'adhésion, dont la négociation est réservée aux seuls acteurs structurants. Ces derniers sont en capacité de définir en

²⁰² C. Doctorow, propos recueillis par H. Guillaud, « Dé-monopoliser l'Internet par l'interopérabilité », *InternetActu*, 11 octobre 2021.

²⁰³ PeREN, Rapport préc., p.1.

²⁰⁴ Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022.

toute indépendance les exigences requises pour parvenir à une interopérabilité avec leur technologie propriétaire, imposant pour ce faire des modalités et conditions, des exigences techniques et des processus de certification qu'ils régissent de manière unilatérale. Par ailleurs, les acteurs structurants contrôlent intégralement l'expérience liée à leurs propres produits et services dès le tout début de l'interaction avec l'utilisateur, collectent les données de ce dernier qui leur sont utiles et, dans la plupart des cas, sont totalement autonomes dans la résolution des questions techniques, dictant les performances et processus techniques aux entreprises utilisatrices tierces. Lorsque l'assistant vocal est configuré par défaut, lié au système d'exploitation, il fait obstacle à l'installation d'un autre assistant vocal sur un terminal. En outre, certains assistants vocaux ne peuvent être configurés pour changer de service de recherche, les seules sources interrogées pour répondre aux questions de l'utilisateur sont présélectionnées par le terminal. Enfin, dans le cadre spécifique des enceintes connectées, les développeurs d'applications pour ces enceintes ne peuvent interfacer leurs services avec ceux de l'enceinte, mais sont conditionnés à fournir des applications en passant par l'assistant vocal lequel ne peut les recevoir que si elles sont intégrées à leur système²⁰⁵.

Il en résulte, à l'échelle des API, que seulement quelques acteurs d'assistants vocaux polyvalents peuvent déterminer qui peut fonctionner, avec quels services et quels appareils. Agir sur l'interopérabilité des écosystèmes permet d'éviter la fragmentation du marché conduisant les entreprises utilisatrices tierces à développer une pluralité de contenus et de fonctionnalités de manière à être accessibles via plusieurs assistants vocaux, et par conséquent être présentes (ou pas) sur chaque écosystème. En théorie, cela permettrait de promouvoir la liberté de choix des utilisateurs (en matière de qualité, de prix ou autres éléments tels que la sécurité ou la confidentialité des services utilisés) et donc l'innovation (de nouveaux services pourraient plus facilement se créer et accéder aux plateformes nécessaires pour être distribués). En revanche, le coût de déploiement de la fonctionnalité et de l'accessibilité est tellement élevé que cela pourrait conduire à remettre en cause le modèle d'affaires des entreprises utilisatrices condamnées à y procéder ou à disparaître du marché. Sont visés les acteurs structurants (Apple, Amazon, Google...) laissant dans l'ombre des milliers d'entreprises utilisatrices qui peinent à se hisser sur le marché. D'autant que, pour l'utilisateur, le choix d'un assistant vocal se porte en règle générale avant tout sur la réputation, la marque. L'accès aux contenus intervient dans un deuxième temps, et en raison de l'absence ou l'incomplétude de l'interopérabilité entre les systèmes, l'utilisateur se trouve rapidement prisonnier d'un système intégré, difficile à paramétrer, et pour lequel le désabonnement reste une opération complexe.

Les propositions pour en sortir. Le principe de bon sens en la matière est de concentrer l'essentiel des obligations sur les plateformes structurantes de manière à ce qu'elles se rendent interopérables avec les entreprises utilisatrices tierces.

²⁰⁵ Commission européenne, Rapport préc., point 380.

Approuvant la décision de la Commission ayant condamné Microsoft pour le refus, constitutif d'un abus de position dominante, de fournir les informations relatives à l'interopérabilité du système d'exploitation Windows avec des produits concurrents, le Tribunal de première instance a ainsi observé que, s'il est établi que les informations ont un caractère indispensable pour Microsoft, elles sont nécessairement de grande valeur pour les concurrents qui souhaitent y avoir accès²⁰⁶. A l'argument, considérant que la divulgation des protocoles de serveur à serveur était soumise à des droits de propriété intellectuelle ce qui impliquait qu'ils devaient rester secrets afin de ne pas porter atteinte à l'innovation ce que Microsoft justifiait par l'exception de libre concurrence établie par les arrêts Magill²⁰⁷ et IMS Health²⁰⁸, le Tribunal répond qu'il n'existe aucune justification objective ou des circonstances exceptionnelles autorisant le refus d'octroyer une licence au sens des jurisprudences précitées²⁰⁹. Il en résulte que l'exigence d'interopérabilité s'impose à Microsoft pour préserver la libre concurrence dans le respect des droits des utilisateurs.

Aujourd'hui, la difficulté dans ce paysage fragmenté réside dans le fait de prendre en considération les outils préexistants pour construire l'interopérabilité. Les données détenues par les acteurs structurants doivent circuler sans entrave technique - que l'on ne pourrait justifier par des questions de sécurité ou d'intérêts juridiquement protégés - et être mises au service de la concurrence effective et de la protection des utilisateurs. A ce titre, dans l'affaire Nintendo, la CJUE a questionné les mesures techniques de protection qui pourraient aller au-delà de l'objectif de lutte contre la production de copies contrefaisantes de jeux vidéo pour contourner les enjeux d'interopérabilité et segmenter le marché. Au visa du considérant 48 de la directive 2001/29, la Cour a rappelé que « *la protection juridique des MTP doit respecter le principe de proportionnalité et ne doit pas interdire les dispositifs ou activités qui ont, sur le plan commercial, un objet ou une utilisation autre que le contournement de la protection technique* ». Ainsi, même si elle n'a pas pris position sur les mesures techniques de protection du système Nintendo, la CJUE a précisé que « *pour l'examen du respect de l'article 6 de la directive 2001/29, les juridictions nationales, au moment d'apprécier le principe de proportionnalité, doivent tenir compte de la finalité des MTP et de la finalité des dispositifs permettant leur contournement* ». Le choix d'autres mesures techniques de protection « *qui auraient pu causer moins d'interférences avec les activités des tiers ne nécessitant pas l'autorisation du titulaire des droits d'auteur ou moins de limitations de ces activités, tout en apportant une protection comparable pour les droits dudit titulaire* », doit être privilégié²¹⁰. Il convient également de vérifier que les mesures techniques de protection sont proportionnées aux usages.

²⁰⁶ TPI Grande chambre, 17 septembre 2007, Microsoft Corp. Contre Commission, aff. T-201/04, point 694.

²⁰⁷ CJCE, 6 avril 1995, RTE et ITP/Commission (« Magill »), aff. C-241/91 P et C-242/91 P.

²⁰⁸ CJUE, IMS Health/NDC Health, aff. C-418/01, 29 avril 2004.

²⁰⁹ TPI, Microsoft, préc., point 680.

²¹⁰ CJUE, 23 janvier 2014, Nintendo Co. Ltd e.a./PC Box Srl, aff. C-355/12, point 32.

Dès lors, il convient d'évaluer le coût de la mise en place d'un standard afin de ne pas renforcer la position des acteurs structurants, et exclure de nouveaux les entreprises utilisatrices tierces. Le PeREN relève à cet effet que les standards AWS d'Amazon et Azur de Microsoft sont interchangeables et utilisables sur des outils concurrents, mais dans la mesure où ils peuvent modifier les API unilatéralement, les entreprises utilisatrices restent dépendantes de ces acteurs structurants. Le pôle de régulation préconise pour y remédier d'encadrer les évolutions brutales de ces API²¹¹. De son côté, la Commission européenne, dans le cadre de l'enquête sectorielle sur l'IoT grand public, constate la complexité du paysage de normalisation couplée à la fragmentation toute aussi importante du paysage des technologies propriétaires, affectant la croissance potentielle et l'innovation, en enfermant conséquemment les utilisateurs dans des écosystèmes propriétaires en l'absence de compatibilité et d'interopérabilité entre les produits, services et technologies d'un même fournisseur. Elle propose de renforcer le processus de normalisation et de le simplifier²¹². Le DMA instaure des obligations qui reprennent ces propositions mais en adoptant une approche plus transversale sans aller jusqu'à promouvoir une approche asymétrique qui pourrait se justifier en raison de la position des acteurs structurants sur chaque écosystème.

L'interopérabilité verticale promue par le DMA. La contestabilité des marchés suppose dans le cadre d'une approche par la régulation *ex ante*, que l'on pose des obligations à l'endroit des acteurs structurants de manière à garantir l'interopérabilité des plateformes.

Les obligations générales en la matière sont posées à l'article 5 du texte. Cependant, le DMA introduit plus spécifiquement, des obligations d'interopérabilité à la fois horizontales à l'égard des services de messagerie (article 7) mais également verticales inscrivant une obligation d'accès aux fonctionnalités essentielles des systèmes d'exploitation, assistants virtuels ou aux capacités matérielles d'un appareil donné que les acteurs structurants utilisent pour leurs propres produits ou services (article 6.7) et la possibilité d'installer des magasins d'applications tiers ou de télécharger des applications sans passer par le magasin d'applications de l'acteur structurant (article 6.4). Seules les obligations d'interopérabilité verticale sont applicables aux assistants vocaux afin de faciliter l'entrée des entreprises utilisatrices tierces leur donnant un accès à des fonctionnalités essentielles qu'elles ne pourraient reproduire, ce qui peut avoir pour certaines, un effet levier pour s'étendre sur le marché. Enfin, l'article 6.9 qui s'appliquera aussi aux assistants virtuels dispose à l'égard de la portabilité dynamique des données des utilisateurs (donc de l'interopérabilité : « *y compris en fournissant gratuitement des outils facilitant l'exercice effectif de cette portabilité des données, et notamment en octroyant un accès continu et en temps réel à ces données.* »). Cela pourrait en théorie dans ce contexte faciliter le passage d'un service d'assistant virtuel à un autre.

²¹¹ PeREN, Rapport préc. p.7.

²¹² Commission européenne, Rapport préc., points 496 et suivants.

Les développeurs d'applications tiers pourront utiliser tous les assistants vocaux à des coûts moins élevés, en choisir certains à tout le moins ceux adaptés à leurs segments d'activités. Cependant le DMA joue sur les deux versants, en envisageant les restrictions qui pèsent sur l'utilisateur, lequel doit être en capacité « *d'installer et d'utiliser effectivement des applications logicielles ou des magasins d'applications logicielles de tiers sur le matériel ou les systèmes d'exploitation d'un contrôleur d'accès (...)* » (considérant 50). Si de telles restrictions ne sont pas levées, c'est la contestabilité des services et produits des acteurs structurants qui fera défaut en s'opposant à ce que les utilisateurs puissent changer d'assistant vocal (considérant 54). Les textes limitent le dispositif aux fournisseurs de services ou de matériels concurrents qui, pour garantir la liberté de choix de l'utilisateur, ont besoin de cet accès (considérant 55) car il s'agit des acteurs clés de l'accès au marché des assistants vocaux. C'est la raison pour laquelle l'acteur structurant doit détailler les conditions techniques afin d'accéder au système d'exploitation (fonctionnalités, sécurité et intégrité du système, modification des interfaces) et au besoin préciser le coût de l'accès au système d'exploitation. Il s'agit de permettre, en aval, aux entreprises utilisatrices tierces de s'interconnecter à l'aide d'interfaces semblables à celles des acteurs structurants, aux services et matériels qu'ils proposent. Il n'en demeure pas moins que la question d'accès est entendue largement et suppose que l'on fournisse aux entreprises utilisatrices tierces des conditions équivalentes à celles prévues pour le contrôleur d'accès. Ce qui ne va pas manquer d'introduire des difficultés en l'absence de standard. Le considérant 96 reconnaît que la mise en œuvre de certaines obligations telles que celle relative à l'interopérabilité pourrait être facilitée par l'utilisation de normes techniques. Cependant, même si le développement d'une norme d'interface ouverte pourrait constituer une alternative, cela reste lourd et complexe en l'absence de consensus.

Il est donc préférable de laisser les acteurs structurants gérer les interfaces en leur permettant de garantir l'intégrité de leurs interfaces comme le prévoit l'article 6.7, par l'octroi de licences à l'endroit des demandeurs d'accès²¹³. A l'égard des utilisateurs, l'acteur structurant pourra prendre des mesures afin qu'ils soient enclins à protéger la sécurité de leurs applications, à la seule condition qu'il ne s'agisse ni d'une pré-installation, ni d'un paramétrage par défaut. Qu'il s'agisse des mesures techniques ou contractuelles, elles doivent rester « *strictement nécessaires et proportionnées et (être) dûment justifiées par le contrôleur d'accès* »²¹⁴. L'acteur structurant peut octroyer des licences sans qu'elles obéissent à des critères objectifs posés préalablement, le DMA ne posant aucune obligation en ce sens. Cependant, si l'acteur structurant refuse le bénéfice de la licence, une procédure pourra être initiée à l'échelle européenne à partir des plaintes des entreprises utilisatrices tierces ou si la Commission s'autosaisie. Ces plaintes sont d'ailleurs complexes à circonscrire *ab initio* en raison des multiples cas de figure pouvant se présenter d'autant que l'acteur structurant est en droit de prendre les mesures nécessaires visant à garantir que les

²¹³ M. Bourreau, « DMA : obligations d'interopérabilité horizontale et verticale », CERRE, novembre 2022, p.14 et s.

²¹⁴ DMA, considérant 50 - v. également les articles 6.4, 6.7.

conditions d'octroi de la licence ne compromettent pas l'intégrité et la sécurité du système. Dès lors, l'octroi d'une licence d'accès reposant sur des critères objectifs pourrait être considérée comme une mesure nécessaire et proportionnée²¹⁵. Pour garantir l'innovation et une concurrence effective qui préservent les fonctionnalités de l'assistant vocal d'acteurs alternatifs économiquement fragiles, certains relèvent toutefois à cet égard que les coûts d'accès aux services et matériels devront être couverts au moins partiellement par les entreprises utilisatrices tierces²¹⁶. Or, le DMA instaure en son article 6, la gratuité de l'accès, ce qui ne manquera pas d'occasionner des différends si cette disposition présente un effet contreproductif sur le maintien de la concurrence.

Il serait probablement nécessaire d'aller plus loin dans la démarche en créant une procédure de règlement des différends portées par un régulateur disposant de compétences techniques suffisantes pour traiter des questions d'interopérabilité et garantir l'intégrité de l'assistant vocal, à l'image de ce que réalise l'ARCEP dans le cadre des communications électroniques²¹⁷. L'article 8 du DMA prévoit à cet effet que les obligations spécifiques pourront être précisées si nécessaire dans le cadre d'actes délégués.

C - L'accès aux données (pour ouvrir le marché)

Le marché des assistants vocaux est dominé par un petit nombre de fournisseurs très puissants qui contrôlent à la fois l'accès aux services et produits proposés aux utilisateurs, mais également aux données qui y sont liées. La massification des bases de données des acteurs structurants (1) génère de la part de tous les nouveaux entrants des inquiétudes légitimes sur la récupération des données utiles aux développeurs d'applications afin d'enrichir l'offre de contenus ou aux éditeurs de contenus pour s'adapter au profil des utilisateurs (2). Cette récupération de données peut se réaliser seulement si l'interopérabilité est garantie, puisqu'il s'agit ici de l'un des leviers à mettre en œuvre (v. *supra*) dans le cadre d'une portabilité des données (3).

1. La massification des bases de données des acteurs structurants

Masse de données. Dans une enquête consacrée aux marchés numériques actualisée en juillet 2022, la House of Representatives identifie comme principal enjeu

²¹⁵ M. Bourreau, « DMA : obligations d'interopérabilité horizontale et verticale », préc.

²¹⁶ *Ibid.*

²¹⁷ V. également sur ce 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* ».

du marché des assistants vocaux la collecte massive de données. Il s'agit d'entraîner et valider les systèmes algorithmiques : « *L'échelle des utilisateurs générant des données est sans doute l'atout le plus important en termes d'IA. Les opérateurs historiques ont accès à de vastes ensembles de données qui, lorsqu'ils sont combinés à l'apprentissage automatique et à l'IA, leur permettent de bénéficier d'économies d'échelle dans le domaine de la maison intelligente* »²¹⁸. Dans le cadre de l'assistant vocal, la masse de données permet d'améliorer les performances des algorithmes d'interaction vocale (Speech-to-Text et Text-to-Speech) et de traitement automatique de langage naturel.

Stockage dans le cloud. Ensuite, le rapport de l'enquête revient sur les investissements importants réalisés par les acteurs structurants pour bâtir l'univers de l'assistant vocal à la fois pour acquérir l'infrastructure physique mais également pour développer les fonctionnalités logicielles dans le cadre d'une infrastructure en nuage: « *La concurrence en tant que plateforme d'assistant vocal nécessite également des ressources financières importantes. Une entreprise doit consentir des investissements importants pour concevoir et former un assistant vocal, ainsi que pour acquérir l'infrastructure physique : matériel et informatique en nuage. En outre, les opérateurs historiques ont également acquis diverses entreprises spécialisées dans la reconnaissance vocale et le traitement du langage naturel, une fonctionnalité utilisée dans leurs assistants vocaux. Par exemple, Apple et Amazon ont tous deux acquis des entreprises pour développer leurs technologies de base en matière de reconnaissance vocale, et tous les opérateurs historiques ont continuellement investi dans des start-ups spécialisées dans l'IA pour améliorer leur écosystème d'assistants vocaux. À l'heure actuelle, les logiciels d'assistants vocaux sont construits sur une infrastructure en nuage. Dans le cas d'Amazon Alexa et de Google Assistant, les plateformes d'assistants vocaux possèdent également l'infrastructure de cloud computing sous-jacente, AWS et GCP, respectivement. Les acteurs du marché notent que les progrès des écosystèmes d'assistants vocaux commencent à s'appuyer sur la technologie de l'edge computing, qui rapproche le calcul et le stockage des données de l'appareil et constitue une technologie dans laquelle les leaders du marché du cloud ont une longueur d'avance* »²¹⁹.

Collecte massive. Enfin, le rapport insiste sur le fait que les investissements matériels réalisés sont peu rentables et l'objectif n'est pas qu'ils le soient. Ils se déploient à grande échelle et à des conditions attractives pour privilégier la collecte massive de données auprès des utilisateurs : « *Les acteurs du marché s'inquiètent également du fait que les entreprises en place offrent du matériel adapté à la voix - en particulier des concentrateurs tels que des enceintes intelligentes - pour collecter de grandes quantités de données personnelles des utilisateurs et renforcer d'autres secteurs d'activité. Patrick Spence, PDG de Sonos, explique : Google et Amazon ont inondé le*

²¹⁸ House of Representatives, *Investigation of competition in Digital Markets*, majority staff reports and recommendations, Part I, juillet 2022, p. 101&s.

²¹⁹ *Ibid.*

marché de produits dont le prix est considérablement subventionné. En effet, ils ne font pas semblant de savoir que les produits eux-mêmes sont déficitaires et ils les offrent régulièrement à des prix très bas, voire gratuitement. Il est difficile de prédire l'impact que les assistants vocaux auront sur la recherche et le commerce électronique, mais les haut-parleurs activés par la voix ont le potentiel de modifier radicalement la façon dont les consommateurs interagissent avec Internet. Nous pensons que Google et Amazon ont été prêts à renoncer à des bénéfices dans les haut-parleurs intelligents pour cette raison, en plus de leur capacité à monétiser les précieuses données domestiques que ces produits aspirent. Et si l'achat et la recherche vocaux deviennent la prochaine grande nouveauté, ils posséderont le marché parce que leur stratégie réussit. Ces deux entreprises contrôlent maintenant environ 85 % du marché américain des enceintes intelligentes. Ce n'est pas parce que leurs activités matérielles sont rentables en soi »²²⁰.

Colonisation des écosystèmes par la donnée. Cette approche de l'assistant vocal illustre le phénomène de « colonisation des écosystèmes par la donnée », ce que constate également la Commission européenne dans le cadre de son enquête sectorielle sur l'Internet des objets. Elle souligne en effet que les entreprises utilisatrices ne disposent pas « *d'un accès cohérent et immédiat aux données pertinentes sur l'utilisation de leurs services et appareils intelligents sur les assistants vocaux et systèmes d'exploitation d'appareils intelligents tiers (...). L'absence de collecte automatique de données empêche les fournisseurs tiers de services de personnaliser l'expérience de l'utilisateur en temps réel et les place dans une situation de désavantage concurrentiel par rapport aux services de premier niveau offerts par les principaux fournisseurs d'assistants vocaux polyvalents* »²²¹.

2. Les données pertinentes

La donnée. Les entreprises qui ne disposent pas d'un accès à la donnée sont largement dépendantes des acteurs qui le sont ; ceux qui sont en capacité de croiser des données des utilisateurs et des services disponibles sur l'assistant vocal sont les seuls en capacité de concurrencer les autres acteurs. Autrement dit, seuls les acteurs structurants peuvent concurrencer d'autres acteurs structurants, car ils sont les seuls à disposer d'une vision systémique²²² de tous les services composant les écosystèmes, en se saisissant de la donnée comme le système sanguin se saisit du corps humain. Ce n'est pas un hasard si le DMA définit la donnée²²³, ce que le Règlement (UE) 2016/679 n'avait pas réalisé en cherchant à créer des catégories de données. Par-delà, l'importance tient à la donnée en elle-même, qu'elle soit

²²⁰ *Ibid.*

²²¹ Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022, point 483.

²²² V. D. Bosco, « A propos du concept de contrôleur d'accès dans le DMA », *Concurrences*, n°3, 2022, p. 64.

²²³ DMA, article 2, « données » : « toute représentation numérique d'actes, de faits ou d'informations et toute compilation de ces actes, faits ou informations, notamment sous la forme d'enregistrements sonores, visuels ou audiovisuels ».

personnelle ou non personnelle. L'écosystème s'accommode d'une vision globale de la donnée.

Les données collectées. Les plateformes numériques, dont celles des assistants vocaux, concentrent les données notamment pour améliorer le fonctionnement des assistants vocaux ou à des fins commerciales. On peut identifier à cet effet particulièrement trois catégories que l'on classe soit parmi les données à caractère personnel, soit parmi les données non personnelles. Tout d'abord, sont concernées les données de validation et d'entraînement des algorithmes, prélevées à partir des fouilles de textes, des bases de données, des profils d'utilisateurs, etc. Ensuite, sont collectées les données d'usage regroupant données à caractère personnel, données contextuelles et autres métadonnées dans l'objectif est de cibler au mieux les besoins de l'utilisateur, possiblement à des fins publicitaires. A ce titre, les données à caractère personnel collectées sont nombreuses comprenant les coordonnées de l'utilisateur (nom, adresse électronique, informations de localisation, adresse IP) et des données relatives à sa personnalité, sa profession voire ses coordonnées bancaires. Enfin, les données de performance de l'assistant vocal sont valorisées pour en améliorer son fonctionnement et promouvoir l'innovation. Ces données, comme les précédentes, comprennent à la fois des données à caractère personnel, et notamment les requêtes des utilisateurs, mais encore des métadonnées et données de contexte, qui dévoilent par exemple l'aboutissement ou non de la requête ou alors le nombre de requêtes formulées par l'utilisateur pour l'écoute d'un même extrait musical. Sont recueillies plus globalement les données relatives à l'utilisateur, que ce soit sous la forme d'informations saisies par celui-ci ou d'informations sur son comportement recueillies dans le cadre du fonctionnement de l'appareil ou de l'accès au service.

Le contrôle des données pour un meilleur contrôle des marchés. En réalité, la difficulté d'accès aux données par les entreprises utilisatrices tierces résulte de la maîtrise du processus technique par l'acteur structurant : allant de la maîtrise de l'infrastructure de cloud dans laquelle les données sont stockées, du développement à l'échelle des systèmes algorithmiques, en passant par le processus d'intégration des applications vocales, l'assistance technique, jusqu'à la création du compte utilisateur qui permet de contrôler l'expérience avec l'utilisateur. Il n'existe pas « d'interopérabilité sémantique » autorisant une certaine forme de standardisation des formats et des modèles de données, une redistribution verticale des données vers les entreprises utilisatrices qui pourraient ainsi traiter de leurs données de performance et d'usage, en lien avec leur propre base de données et en utilisant leurs propres outils²²⁴. Il convient donc de la promouvoir.

Si l'accès aux données permet au fournisseur d'assistant vocal de se déployer, l'absence d'accès à ces mêmes données crée des obstacles à l'entrée de nouveaux arrivants sur le marché et au développement des concurrents. Les répondants à

²²⁴ PeREN, *Eclairage sur : l'interopérabilité*, octobre 2021, p. 4.

l'enquête sectorielle de la Commission sur les objets connectés grand public soulignent à cet effet que les fournisseurs d'assistants vocaux « *ont un accès complet aux données des consommateurs, tandis que les développeurs tiers doivent s'occuper d'une procédure de configuration vraiment compliquée avant que l'utilisateur puisse nous donner ses données, par exemple sa localisation, pour recevoir des informations plus pertinentes* »²²⁵. Au-delà, les acteurs structurants peuvent être en mesure de dégrader l'utilisation des services et applications proposées par les entreprises utilisatrices tierces en ne faisant pas aboutir les requêtes de l'utilisateur. Dans le même temps, les acteurs structurants demandent aux entreprises utilisatrices tierces, et plus particulièrement aux développeurs d'applications, de leur transmettre une quantité importante de métadonnées et notamment dans le cadre des procédures d'intégration sur la plateforme de l'assistant vocal. Or, les développeurs d'applications n'ont pas de visibilité sur l'utilisation des données par les acteurs structurants et l'aboutissement des requêtes de l'utilisateur. Pour améliorer la qualité de leur service, il leur est nécessaire d'accéder tant aux données de performance qu'aux données d'usage.

Accès aux données d'usage. En premier lieu, le fournisseur d'assistant vocal ne devrait pas être en capacité de capter toutes ces données sans laisser les fournisseurs d'applications tiers en disposer pour gérer l'expérience avec l'utilisateur. A cet égard, l'accès aux données d'usage est un enjeu essentiel. Un premier pas avait été fait en ce sens à l'occasion de l'adoption du Règlement 2019/1150. L'article 9 de ce texte demeure en réalité bien en deçà de la reconnaissance d'un droit d'accès à ces données puisqu'il ne consacre qu'une simple obligation de transparence à la charge des intermédiaires de plateformes sous la forme d'une obligation d'information - à faire figurer dans leurs CGU - sur l'option qu'ils sont libres d'offrir à leurs utilisateurs professionnels quant à l'accès à ces données d'usage. Le DMA va bien plus loin et consacre explicitement cet accès aux données d'usage en faveur d'entreprises tierces. Ainsi, son article 6.10 exige que « *le contrôleur d'accès assure gratuitement aux entreprises utilisatrices et aux tiers autorisés par les entreprises utilisatrices, à leur demande, un accès et une utilisation effectifs, de haute qualité, continus et en temps réel en ce qui concerne les données agrégées et non agrégées, y compris les données à caractère personnel, fournies ou générées dans le cadre de l'utilisation des services de plateforme essentiels concernés ou des services fournis conjointement aux services de plateforme essentiels concernés, ou à l'appui de ceux-ci, par ces entreprises utilisatrices et par les utilisateurs finaux qui se servent des produits et services fournis par ces entreprises utilisatrices* ». S'agissant des données à caractère personnel, il précise que « *le contrôleur d'accès ne donne un tel accès aux données à caractère personnel et ne les utilise que lorsque les données sont directement liées à l'utilisation faite par les utilisateurs finaux en lien avec les produits ou services que l'entreprise utilisatrice concernée fournit par l'intermédiaire du service de plateforme essentiel concerné, et lorsque les utilisateurs finaux optent pour un tel partage de données en donnant leur consentement* ». Les modalités d'accès aux données seront

²²⁵ Commission européenne, Enquête préc., point 483.

alors déterminantes. A cette fin, le considérant 60 précise que « *les contrôleurs d'accès devraient [...] garantir l'accès continu et en temps réel à de telles données au moyen de mesures techniques appropriées, par exemple en mettant en place des interfaces de programmation de haute qualité ou des outils intégrés pour les entreprises utilisatrices de petit volume* ».

Accès aux données de performance ? S'agissant des données de performance, l'article 5.8 du DMA prévoit que « *le contrôleur d'accès fournit aux annonceurs et aux éditeurs, ainsi qu'aux tiers autorisés par les annonceurs et les éditeurs, à leur demande et gratuitement, un accès aux outils de mesure de performance du contrôleur d'accès et aux données qui leur sont nécessaires pour effectuer leur propre vérification indépendante de l'inventaire publicitaire, notamment les données agrégées et non agrégées. Ces données sont fournies de manière à permettre aux annonceurs et aux éditeurs d'utiliser leurs propres outils de vérification et de mesure afin d'évaluer la performance des services de plateforme essentiels fournis par le contrôleur d'accès* ». Cette disposition vise la fourniture de service de publicité en ligne. Néanmoins, garantir la contestabilité et l'équité des marchés numériques suivant l'objectif du DMA pourrait conduire à considérer la nécessité de consacrer plus largement un droit d'accès aux données de performance au profit des fournisseurs d'applications tiers afin de leur permettre, lorsque leur application est accessible via un assistant vocal, d'améliorer la qualité de la réponse à la question posée par l'utilisateur et par conséquent la pertinence du classement de leur application. Cela permettrait en outre d'évaluer les cas où la requête utilisateur aurait été mal comprise par l'assistant vocal et dans lesquels celui-ci jouerait alors, par défaut, l'application de l'entreprise qui est dans le même temps le fournisseur de l'interface.

Renouer une relation directe avec l'utilisateur. Au-delà, l'enjeu consiste à permettre aux fournisseurs d'applications de renouer une relation plus directe avec l'utilisateur abonné à leurs services, ce qui pourrait être une autre façon d'accéder à ces données d'usage. Tel est le sens de l'article 5.5 du DMA selon lequel « *le contrôleur d'accès permet aux utilisateurs finaux, par l'intermédiaire de ses services de plateforme essentiels, d'accéder à des contenus, abonnements, fonctionnalités ou autres éléments et de les utiliser en se servant de l'application logicielle de l'entreprise utilisatrice, y compris lorsque ces utilisateurs finaux ont acquis de tels éléments auprès des entreprises utilisatrices concernées sans avoir recours aux services de plateforme essentiels du contrôleur d'accès* ».

3. La portabilité des données

La question de la portabilité des données est souvent traitée sous l'angle des utilisateurs (v. infra), mais elle présente un volet tout aussi intéressant qui se matérialise par la relation existante entre l'assistant vocal et les entreprises

utilisatrices tierces dans la perspective de faire migrer vers ces dernières les données centralisées par l'assistant vocal. Idéalement, les fournisseurs d'assistants vocaux devraient fournir aux entreprises utilisatrices tierces toutes les données collectées en relation avec leur service en continue et gratuitement. Pour l'heure, elles utilisent les données en question pour développer des services concurrents.

La portabilité des données doit s'organiser dans le cadre de formats standardisés. A défaut, le PeREN précise qu'elle pourrait se réaliser « à l'aide de connecteurs accessibles à travers des API »²²⁶. Cette approche est déjà développée entre acteurs structurants (v. par exemple le *Data Transfer project*) et pour interconnecter certains éléments de services en lignes (IFTT)²²⁷. Dès lors, la portabilité des données de l'utilisateur d'une plateforme vers une autre est envisageable, et ce de manière automatique lors de l'inscription de l'utilisateur ce qui permet de réduire les coûts de migration. L'objectif est de permettre la décentralisation de la donnée, en exigeant des acteurs structurants « une portabilité asymétrique », lesquels se verraient contraints de garantir les différents aspects liés à la sécurité des données, comme le consentement de l'utilisateur, la confidentialité des données à caractère personnel et les règles de sécurité attachées à la protection des consommateurs. Ces initiatives sont encore réservées à la migration de données entre les seuls acteurs structurants, mais ne semblent pas encore visées les entreprises utilisatrices probablement en raison des coûts de migration²²⁸. Dans la mesure où la portabilité des données répond à un objectif d'interopérabilité, le respect de cette obligation peut se révéler essentiel au titre d'une approche par la régulation *ex ante*.

Dans le contexte, la portabilité des données vers les entreprises utilisatrices fait état d'une difficulté supplémentaire dans la mesure où l'agrégation peut à la fois concerner des données à caractère personnel et non personnelles. Dès lors, elles n'obéissent pas aux mêmes règles. L'utilisateur doit consentir à la portabilité de ses données personnelles auprès de l'entreprise utilisatrice tierce. Les conditions d'obtention du consentement sont encore à préciser, car il peut être requis via l'interface de l'assistant vocal, ou par les entreprises utilisatrices²²⁹.

En interdisant l'autopréférence pour ne pas bloquer l'accès au marché, en consacrant l'interopérabilité pour le favoriser ou encore en garantissant l'accès aux données pour ouvrir le marché, les réformes du droit de l'Union européenne pourraient venir au soutien du développement des acteurs du secteur culturel qui ont désormais recours aux assistants vocaux pour diffuser leurs offres. Cette action favorise dans le même temps un accès aux contenus, gage de diversité.

²²⁶ PeREN, *Eclairage sur : l'interopérabilité*, préc., p.2.

²²⁷ *Ibid.*

²²⁸ *Ibid.*

²²⁹ J. Krämer, *Accès aux données*, CERRE, novembre 2022, p.12.

III - ASSISTANTS VOCAUX ET ACCÈS AUX CONTENUS CULTURELS

Ces nouvelles formes d'interaction humain-machine emportent d'importantes conséquences sur la diffusion et la diversité des contenus culturels. Cela suppose de penser les instruments afin de garantir la diversité au sein de ces écosystèmes.

Afin que l'utilisateur exerce une liberté de choix, il convient d'interroger les pratiques amplifiées par l'usage d'assistants vocaux ou propres à celui-ci en ce qu'elles viennent altérer l'accès aux contenus pour l'utilisateur, particulièrement s'agissant du mécanisme de la réponse unique qui ne lui présentera qu'une seule occurrence. L'étude réalisée par le CSA et l'HADOPI sur les enceintes connectées constatait déjà à cet égard un risque d'enfermement tout en reconnaissant le regain d'intérêt de l'utilisateur pour les plateformes culturelles et les médias sonores permettant une meilleure exposition des contenus²³⁰. Le rapport sur la neutralité des terminaux de l'ARCEP²³¹ insistait également sur ce point en soulignant la nécessité de garantir la liberté de choix de l'utilisateur pour lui permettre de faire des choix éclairés. En effet, les assistants vocaux proposent « *une expérience intelligente mais orientée de l'Internet* », avec le risque de voir les utilisateurs « *enfermés* » dans des univers de marques distribuant à la fois les terminaux, les contenus et éventuellement les services. Le fonctionnement de ces appareils repose également sur une collecte importante de données personnelles, par le simple enregistrement des requêtes vocales comme par l'analyse du contenu de ces requêtes. Ainsi, la liberté de choix de l'utilisateur supposera que la question du consentement soit adaptée à l'interface vocale, pour une meilleure information mais aussi pour éviter une « *monétisation de l'intime* »²³². De même, dans le cadre de son enquête sectorielle sur les objets connectés grand public, la Commission européenne a exposé ses préoccupations concernant des pratiques qui pourraient avoir un impact sur la concurrence, l'innovation et le choix des utilisateurs²³³. En précisant qu'en bout de chaîne, il en va du nombre d'applications vocales mises à la disposition de l'utilisateur, conditionné par des enjeux d'interopérabilité, d'installation par défaut de ces applications, du paramétrage, de l'exclusivité, de la désintermédiation. Comme cela a été souligné, le caractère intuitif des interfaces utilisateurs masque de nombreux paramètres déterministes dans le calcul des résultats de requêtes se limitant souvent à un seul résultat dont le choix n'est jamais neutre²³⁴.

²³⁰ HADOPI / CSA, *Assistants vocaux et enceintes connectées*, 28 mai 2019.

²³¹ ARCEP, *Smartphones, tablettes, assistants vocaux... Les terminaux, maillon faible de l'ouverture d'Internet : Rapport sur leurs limites et les actions à envisager*, février 2018.

²³² CNIL, *A votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux*, Livre blanc n°1, septembre 2020.

²³³ Commission européenne, *Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs*, 20 janvier 2022 ; v. également House of Representatives, *Investigation of competition in Digital Markets*, majority staff reports and recommendations, Part I, juillet 2022, p. 101&s.

²³⁴ A. Masure, propos recueillis par H. Guillaud, Vox Machines, « Si les assistants vocaux sont la solution, quel est le problème ? », *InternetActu*, 19 décembre 2018.

Cela suppose d'agir sur différents leviers, selon une démarche en trois temps. De prime abord, il paraît nécessaire de lever les restrictions qui pèsent sur le marché en questionnant les pratiques d'autopréférence des opérateurs d'assistants vocaux privilégiant leurs propres services, par exemple lorsque l'assistant vocal répond à la requête vocale de l'utilisateur en privilégiant ses services connexes (A). Au-delà de l'accès garanti à une pluralité de contenus, la liberté de choix suppose en outre d'assurer la diversité de ces contenus. Il s'agit notamment d'envisager l'intégration technique des applications existantes pour garantir la diversité. Cela permet alors non seulement de garantir l'accès à plusieurs contenus, ce que comprend le pluralisme, mais aussi de garantir l'accès à des contenus différents (B). Cette approche de la diversité par l'offre doit par ailleurs être complétée par une approche de la diversité par la demande. Pour ce faire, il convient de penser le rôle de l'utilisateur autrement, en lui conférant le moyen d'être informé mais aussi d'agir sur l'accès au contenu, notamment en lui conférant un droit au paramétrage pour élargir son choix de contenus dans l'écosystème, mais également de choisir cet écosystème en levant les barrières à la sortie, ce qui suppose de faciliter ses conditions de désabonnement au service et de lui garantir un droit à la portabilité, et donc l'interopérabilité (C). Cela revient par conséquent à penser les mécanismes de co-régulation de la diversité par l'offre et la demande.

A - Lutter contre l'autopréférence par l'expression du pluralisme : premier levier pour garantir la liberté de choix

L'autopréférence, qui relève en principe de la relation existant entre les acteurs économiques du marché, est ici appréhendée sous l'angle de l'utilisateur. En effet, les mécanismes d'autopréférence utilisés par les acteurs structurants ont une incidence à la fois sur le nombre d'applications vocales mises à sa disposition, mais également sur l'origine de ces applications vocales restreignant par essence la liberté de choix de l'utilisateur. Pas besoin de verrouiller l'écosystème pour que la question de la liberté de choix puisse être analysée sous cet angle ; il suffit simplement que l'acteur structurant ferre l'utilisateur autour de ses produits et services et les propose en priorité à des conditions défiant la concurrence.

Avec l'arrivée du DMA, la question de la diversité des contenus se pose sous l'angle de la contestabilité et du caractère équitable des marchés. Le texte assoit l'une des facettes de la transparence et vise à remédier aux asymétries informationnelles.

Pour promouvoir leurs services, les entreprises utilisatrices se trouvent en effet dans l'obligation de se soumettre aux règles d'intégration posées par les contrôleurs d'accès et leurs services de plateforme essentiels. Quelques magasins d'applications dominant ainsi le marché alors qu'il existe en la matière nombre de magasins d'applications innovants. De la même manière, l'offre d'applications vocales

notamment culturelles peut être conditionnée par l'achat d'un bien de consommation à l'issue d'un mécanisme de croisement de données entre les différents services de plateforme essentiels, particulièrement ceux de l'assistant vocal. Aborder la question de la diversité en partant de la contestabilité de la « fabrique » de la réponse unique constitue probablement le point de départ de la réflexion mais ne fait que produire un résultat quantitatif, reposant sur la multiplicité des contenus mis à disposition, et non sur ce que l'on pourrait concevoir au titre de la diversité.

La diversité, composante de la transparence. La transparence constitue une exigence qui transcende les divisions les plus classiques du droit et des communautés scientifiques. Ce principe doit se reconstruire en permanence pour absorber les nouveaux enjeux et l'innovation. Or, des liens sont désormais à construire entre la transparence et la mise en œuvre d'une diversité par des mécanismes incitatifs. Cette incitation a pu être directe s'agissant de l'application du principe de pluralisme aux anciens modèles de communication²³⁵. Probablement qu'elle se construit de manière plus indirecte par une information permettant de comparer les offres de services et de contenus dès lors que, désormais, la transparence vise également à restaurer la liberté de choix de l'utilisateur dans une logique d'*empowerment*, afin de préserver ses libertés individuelles ainsi que le débat démocratique.

Les enjeux du pluralisme. Il est nécessaire de réfléchir à la manière dont ces mécanismes ont été exploités au siècle dernier pour envisager sous quelle manière ces mécanismes peuvent être adaptés à l'environnement des assistants vocaux pour garantir la liberté de choix de l'utilisateur. Ce lien entre le pluralisme et la liberté de choix a été établi en son temps par le Conseil constitutionnel, pour montrer que sans une pluralité d'acteurs et de contenus, l'utilisateur ne pouvait pas prétendre à l'exercice de sa liberté de choix entre des contenus diversifiés.

Liberté de choix, diversité et pluralisme. Pour que l'utilisateur soit en mesure d'exercer sa liberté de choix, il est nécessaire qu'il ait accès à des contenus diversifiés. Cette diversité doit alors être comprise sous l'angle du pluralisme, érigé d'abord par le droit constitutionnel au titre des fondements de la démocratie, puis par le droit des médias, pour refléter les différents courants de pensée et d'opinion sociaux-culturels. Selon la jurisprudence du Conseil constitutionnel en la matière, le pluralisme doit à la fois garantir la diversité ou à tout le moins la pluralité d'acteurs (pluralisme externe), mais également la diversité ou la pluralité de contenus (pluralisme interne).

En effet, le Conseil constitutionnel et la CEDH font du pluralisme une condition de la démocratie²³⁶, centrée sur la confrontation des idées et des opinions. Le pluralisme procède du rôle primordial joué par la liberté de communication dans le fonctionnement de la démocratie. Par ailleurs, la démocratie ne saurait se limiter à sa

²³⁵ S'agissant de la presse et de l'audiovisuel, v. P. Marcangelo-Leos, *Pluralisme et audiovisuel*, LGDJ, Bibliothèque de droit public, t. 240, 2004.

²³⁶ Cons. const., 18 sept. 1986, n° 86-217 DC ; CEDH, 7 déc. 1976, aff. 5493/72, *Handyside c. Royaume-Uni*.

conception politique ; elle englobe le débat culturel, culturel et social sans lequel l'appréciation du pluralisme serait incomplète. Le pluralisme auquel il est fait référence diffère quelque peu cependant à la lecture des décisions du Conseil constitutionnel et de la CEDH sans pour autant s'opposer. Le pluralisme peut avoir pour signification la pluralité ou la diversité des opinions²³⁷. Cette différence de perception n'est pas sans incidence sur la diversité attendue.

Pour le Conseil constitutionnel, le pluralisme est lié à un ordre de grandeur qui se matérialise en matière de communication audiovisuelle et de presse par la préservation du pluralisme externe (un nombre d'acteurs économiques suffisant pour garantir la diversité). Le pluralisme des contenus, qualifié de pluralisme interne, a seulement été garanti à l'endroit de la communication audiovisuelle. Dans le domaine de la presse, la préservation du pluralisme suppose que les journaux soient en capacité de publier des débats d'intérêt général²³⁸ : les aides distribuées par l'État ont pour objectif de préserver le pluralisme de la presse en identifiant les publications à faibles ressources publicitaires, sans véritablement définir la diversité²³⁹.

Depuis la révision constitutionnelle du 23 juillet 2008, l'article 4 de la Constitution proclame « *les expressions pluralistes des opinions* », mais le Conseil constitutionnel limite l'application de ce principe au pluralisme politique²⁴⁰, retenant à l'égard des autres domaines la qualification d'objectif de valeur constitutionnelle²⁴¹. Cependant l'article 34 de la Constitution a également été révisé à la même date, et commande au législateur de garantir l'indépendance et le pluralisme des médias. La garantie du pluralisme en ressort, du moins en apparence, renforcée car liée à l'indépendance des médias. En réalité, le respect du pluralisme peut s'accommoder de dispositifs techniques propres, même s'ils sont imparfaits, mais l'indépendance ne se décrète pas. D'ailleurs dans un écosystème, la dépendance est la caractéristique première des rapports entre les acteurs, ce qui signifie que la question de l'indépendance se construit dans le cadre d'une approche par la régulation *ex ante*, d'abord avec des critères classiques liés à la taille critique de l'entreprise, et au nombre d'utilisateurs concernés mais à partir desquels il serait possible de rechercher une approche plus fine en « patchant » au fur et à mesure de nouvelles règles²⁴². C'est ainsi que procèdent les textes du DMA (qui instaure une forme de pluralisme externe) et du DSA (plutôt une sorte de pluralisme interne) pour poser leurs obligations à l'égard des plus grandes plateformes, mais sans aller jusqu'à distinguer les obligations en fonction des services et des écosystèmes concernés.

²³⁷ P. Marcangelo-Leos, *Pluralisme et audiovisuel*, oct. 2004, LGDJ, Bibl. dr. public, t. 240 ; « La place du pluralisme dans le droit de la communication audiovisuelle – De la distinction entre pluralisme et pluralité », *Légipresse* juin 2003, n°202.

²³⁸ B. Lamy (de), « La Constitution et la liberté de la presse », *Les Nouveaux Cahiers du Conseil constitutionnel*, n° 36, juin 2012.

²³⁹ Ou par renvoi à des infractions par voie de presse visant à démontrer qu'elles feraient obstacle au pluralisme. Sur fond de valeur sociale, la condamnation sur les cinq dernières années pour des faits d'incitation à la violence et la haine raciale fait obstacle à l'octroi d'une aide : CE, 22 févr. 2017, req. n° 395948, Société Valmonde.

²⁴⁰ Cons. const., 12 janv. 2012, n° 2011-4538 DC.

²⁴¹ Cons. const., 28 mai 2010, n° 2010-3 QPC.

²⁴² V. Weigel, « Pluralisme, anti-concentration et média global », *RLDI* 2008, p. 44.

Enfin, sur ce principe de pluralisme, le Conseil constitutionnel s'est fait l'interprète de l'article 11 DDHC, en consacrant la liberté de recevoir des idées et des opinions. A la lumière de ce texte, il protège le récepteur, imposant aux émetteurs la préservation du caractère pluraliste des courants de pensée et d'opinion sous le contrôle du CSA en matière de communication audiovisuelle et dans le cadre du régime répressif de la presse. De sorte que les récepteurs sont « *à même d'exercer leur libre choix sans que ni les intérêts privés, ni les pouvoirs publics puissent y substituer leurs propres décisions ni qu'on puisse en faire l'objet d'un marché* »²⁴³. Ce considérant du Conseil constitutionnel donne utilement une valeur à la liberté de choix du récepteur dans le cadre de ce que l'émetteur lui propose mais pas dans le cadre de ce qu'il entend choisir. Cette liberté de choix était déjà conditionnelle au XXème siècle, elle l'est davantage désormais donnant au récepteur l'illusion du choix en raison de la question de l'accès aux services, de leur visibilité.

Pluralisme du XXIème siècle. Le principe de pluralisme tel qu'il a été construit par le Conseil constitutionnel est figé en contemplation des pratiques linéaires du XXème siècle, mais n'interdit nullement de réfléchir à l'adaptation de ce principe aux usages numériques, et particulièrement aux assistants vocaux. Le pluralisme doit alors garantir une pluralité d'acteurs sur le marché (1), une diversité de modalités d'accès au contenu (2) ainsi que le pluralisme des contenus (3).

1. Pluralité d'acteurs sur le marché

Deux exemples significatifs sont abordés ici. Le premier met en perspective l'inefficacité des seuils de concentration pour garantir le pluralisme externe face à des acteurs structurants qui passent en dessous des seuils. Le second questionne le respect du pluralisme externe en raison de la remise en cause de la réception du signal hertzien par les voitures nouvelle génération.

Pluralité d'acteurs et pluralisme externe. L'exemple du respect des seuils de concentration. Le contrôle des seuils de concentration des entreprises, mécanisme

²⁴³ Cons. const., 10 et 11 oct. 1984, n° 84-181 DC, Entreprise de presse ; Cons. const., 18 sept. 1986, n° 86-217 DC, Liberté de communication audiovisuelle.

utilisé à la fois par le droit de l'Union²⁴⁴ et le droit interne²⁴⁵ pour garantir le respect du pluralisme externe, interroge l'efficacité de ces seuils pour révéler la place de marché d'un acteur du numérique dans la mesure où le chiffre d'affaires n'est pas en rapport avec l'évaluation de la cible. Cette interrogation était déjà présente au XX^{ème} siècle pour évaluer les effets du contrôle des seuils de concentration sur le pluralisme interne (des contenus). Les seuils capitalistiques posés par la loi du 30 septembre 1986 modifiée ne pouvaient s'apprécier en tant que tels, sans prendre en compte les critères d'audience pour analyser l'impact éditorial d'une entreprise de communication audiovisuelle sur la caractérisation du pluralisme des contenus²⁴⁶. Le cas de l'acquisition de l'application Shazam par Apple illustre l'innocuité de tels seuils lorsqu'une entreprise qui n'entre pas dans le champ d'application du règlement de 2004 est pour autant valorisée à un milliard de dollars. C'est la raison pour laquelle l'Autorité de concurrence de l'Autriche à laquelle se sont jointes six autorités nationales de la concurrence dont la France, a demandé un renvoi à la Commission européenne sur le fondement de l'article 22 du règlement pour que la situation du

²⁴⁴ Article 1er et considérant 20 du Règlement n°139/2004 du 20 janvier 2004 relatif au contrôle des entreprises.

Considérant 20 : (20) « Il est utile de définir la notion de concentration de telle sorte qu'elle couvre les opérations entraînant un changement durable du contrôle des entreprises concernées et donc de la structure du marché. Il convient par conséquent d'inclure dans le champ d'application du présent règlement toutes les entreprises communes accomplissant de manière durable toutes les fonctions d'une entité économique autonome. Il convient en outre de traiter comme une concentration unique des opérations qui sont étroitement liées en ce qu'elles font l'objet d'un lien conditionnel ou prennent la forme d'une série de transactions sur titres effectuées dans un délai raisonnablement bref ».

Article 1er : « Champ d'application : 1. Sans préjudice de l'article 4, paragraphe 5, et de l'article 22, le présent règlement s'applique à toutes les concentrations de dimension communautaire telles qu'elles sont définies au présent article.

2. Une concentration est de dimension communautaire lorsque :

a) le chiffre d'affaires total réalisé sur le plan mondial par l'ensemble des entreprises concernées représente un montant supérieur à 5 milliards d'euros, et

b) le chiffre d'affaires total réalisé individuellement dans la Communauté par au moins deux des entreprises concernées représente un montant supérieur à 250 millions d'euros,

à moins que chacune des entreprises concernées réalise plus des deux tiers de son chiffre d'affaires total dans la Communauté à l'intérieur d'un seul et même État membre.

3. Une concentration qui n'atteint pas les seuils fixés au paragraphe 2 est de dimension communautaire lorsque :

a) le chiffre d'affaires total réalisé sur le plan mondial par l'ensemble des entreprises concernées représente un montant supérieur à 2,5 milliards d'euros ;

b) dans chacun d'au moins trois États membres, le chiffre d'affaires total réalisé par toutes les entreprises concernées est supérieur à 100 millions d'euros ;

c) dans chacun d'au moins trois États membres inclus aux fins du point b), le chiffre d'affaires total réalisé individuellement par au moins deux des entreprises concernées est supérieur à 25 millions d'euros, et

d) le chiffre d'affaires total réalisé individuellement dans la Communauté par au moins deux des entreprises concernées représente un montant supérieur à 100 millions d'euros,

à moins que chacune des entreprises concernées réalise plus des deux tiers de son chiffre d'affaires total dans la Communauté à l'intérieur d'un seul et même État membre.

4. Sur la base des données statistiques susceptibles d'être régulièrement fournies par les États membres, la Commission fait rapport au Conseil sur la mise en œuvre des seuils et critères figurant aux paragraphes 2 et 3 avant le 1er juillet 2009, et peut présenter des propositions conformément au paragraphe 5.

5. À la suite du rapport visé au paragraphe 4 et sur proposition de la Commission, les seuils et les critères mentionnés au paragraphe 3 peuvent être révisés par le Conseil statuant à la majorité qualifiée ».

²⁴⁵ Article L. 430-2 du Code de commerce : « I.- Est soumise aux dispositions des articles L. 430-3 et suivants du présent titre toute opération de concentration, au sens de l'article L.430-1, lorsque sont réunies les trois conditions suivantes :

-le chiffre d'affaires total mondial hors taxes de l'ensemble des entreprises ou groupes de personnes physiques ou morales parties à la concentration est supérieur à 150 millions d'euros ; -le chiffre d'affaires total hors taxes réalisé en France par deux au moins des entreprises ou groupes de personnes physiques ou morales concernés est supérieur à 50 millions d'euros ; - l'opération n'entre pas dans le champ d'application du règlement (CE) n° 139/2004 du Conseil, du 20 janvier 2004, relatif au contrôle des concentrations entre entreprises. II.- Lorsque deux au moins des parties à la concentration exploitent un ou plusieurs magasins de commerce de détail, est soumise aux dispositions des articles L. 430-3 et suivants du présent titre toute opération de concentration, au sens de l'article L. 430-1, lorsque sont réunies les trois conditions suivantes : -le chiffre d'affaires total mondial hors taxes de l'ensemble des entreprises ou groupes de personnes physiques ou morales parties à la concentration est supérieur à 75 millions d'euros ; -le chiffre d'affaires total hors taxes réalisé en France dans le secteur du commerce de détail par deux au moins des entreprises ou groupes de personnes physiques ou morales concernés est supérieur à 15 millions d'euros ; -l'opération n'entre pas dans le champ d'application du règlement (CE) n° 139/2004 du Conseil, du 20 janvier 2004, précité.

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

rachat soit de nouveau examinée, signe d'une réaction de l'Union européenne à la multiplication d'acquisitions d'entreprises innovantes à haute valeur mais affichant un faible chiffre d'affaires²⁴⁷. Se saisissant des éléments soumis par l'Autriche et les Etats se joignant à la demande de renvoi, et sans préjudice de l'issue de son enquête approfondie, la Commission a considéré que l'opération peut avoir un effet négatif significatif sur la concurrence dans l'espace économique européen. La Commission a également conclu qu'elle était l'autorité la mieux placée pour traiter les effets transfrontaliers potentiels de l'opération. La Commission a demandé à Apple de lui notifier l'opération. Pour Philippe Corruble, « *ce phénomène est apparu d'autant plus cruellement que l'acquisition a pu être le fait, dans certaines opérations, d'un acquéreur puissant voire dominant et pour un montant très élevé. Ainsi, des opérations aussi importantes que Facebook/Instagram, Google/Google+, Amazon/Whole Foods ou Google/Waze ont échappé à tout contrôle par la Commission européenne : le règlement ne s'appliquait tout simplement pas. Cette situation appelait de toute évidence une prise de conscience de la nécessité d'un changement d'approche (Cousin, 2019). Elle fut d'abord le fait de certaines ANC* »²⁴⁸. L'article 14 du DMA va obliger les contrôleurs d'accès à informer la Commission de tout projet de concentration, dès lors que les entités qui fusionnent ou la cible de l'acquisition fournissent des services de plateforme essentiels ou tout autre service numérique, ou permettent la collecte de données. Cette obligation va s'imposer dès 2023 indépendamment du fait que l'opération doive être ou non notifiée à la Commission en vertu du règlement concentration, ou à une autorité nationale de la concurrence en vertu des règles nationales, ce qui n'est pas sans rappeler la procédure de l'article 22 du règlement mais avec des critères plus appropriés que ceux qui résultent du règlement de 2004. Les seuils imposés par le droit de la concurrence ne permettent pas de garantir le pluralisme externe. Une approche par la régulation qui met en perspective des critères davantage qualitatifs que quantitatifs est attendue par l'application du DMA (v. *supra*).

Pluralisme externe et signal hertzien. Les assistants vocaux et enceintes connectées ne doivent pas priver les téléspectateurs et auditeurs de la réception du signal hertzien, notamment lorsqu'il s'agit d'un choix de leur part. Il existe encore 20% de récepteurs de services linéaires, et la disparition de la réception hertzienne serait une atteinte au pluralisme externe en ce qu'elle contribuerait à priver le récepteur de sa liberté de choix, même si cette problématique est transitoire. Il n'en demeure pas moins que cette question de « *mix technologique* » pour la réception des services linéaires de radio et de télévision, se joue à l'heure actuelle, notamment pour la radio, car elle dépend en grande partie des équipements des véhicules. Si les constructeurs automobiles décident de ne plus équiper les véhicules d'antenne de réception du signal hertzien, non seulement le récepteur perdra sa liberté de choix et la gratuité de

²⁴⁷ Décision Commission, 6 sept. 2018 Apple/Shazam. M.8788 ; Communiqué de presse de la Commission européenne n° IP/18/664, 06 Février 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.

la réception des services radiophoniques qui seront conditionnés au coût d'un abonnement au réseau voire d'un abonnement aux services, mais encore il sera équipé d'un tableau de bord qui mobilise d'autres difficultés et notamment celle de l'accès aux services de son choix. En maintenant la réception du signal hertzien pour la radio et la télévision, on maintient l'universalité des services proposés, leur gratuité et l'anonymat des choix du récepteur.

2. Diversité de modalités d'accès aux contenus

La diversité des modalités d'accès aux contenus suppose que des points d'attention soient portés aux enjeux de la réponse unique, à l'accès aux terminaux, et plus spécifiquement que l'on interroge, sous cet angle, les interfaces intégrées aux véhicules nouvelle génération.

Pluralisme et réponse unique. La réponse unique fournie par l'assistant vocal doit favoriser la recherche d'outils qui permettraient en amont de la requête de classer des contenus, en raison de leur intérêt général de manière à garantir le débat démocratique. L'objectif à terme serait de proposer des obligations spécifiques soit à l'endroit de l'acteur structurant, soit à l'égard des services fournis par les entreprises utilisatrices.

L'utilisateur, dans le cadre de sa requête, peut ne pas préciser le service qu'il souhaite utiliser pour une action ; l'assistant virtuel proposera alors un service par défaut. Ainsi, pour les flashes d'information, les contenus du service public (Radio France) disponibles, sont aujourd'hui lancés par défaut. Sur Google Assistant et Amazon Echo, les services de radio sont lancés par défaut en passant par l'agrégateur TuneIn. Il appartient à l'utilisateur de modifier cette sélection d'applications vocales dans les paramètres s'il souhaite accéder à d'autres contenus. En pratique, les utilisateurs installent rarement des applications qui offrent la même fonctionnalité qu'une application installée par défaut, sauf à considérer qu'elle ne présente pas la qualité attendue, ou qu'elle suppose un abonnement non souhaité. Dès lors, l'installation par défaut peut être analysée comme répondant aux attentes des utilisateurs qui voudraient utiliser leur assistant vocal dès la prise en main. Par ailleurs, la réponse donnée dépendra toujours de la formulation de la question posée, des classements réalisés par les algorithmes, par les briques logicielles qui mettent des intentions dans la réponse, et du cadre de la réponse (est-elle simplement vocale ?).

Les fabricants d'assistants virtuels n'exigent aucune exclusivité auprès des éditeurs de contenus ; théoriquement ceux-ci peuvent donc être présents dans tous les écosystèmes. Mais certains partenariats sont difficiles à contourner « *car la nature limitée de l'interface de choix du contenu sur les enceintes rend encore plus décisif le*

choix du constructeur »²⁴⁹. De plus, les coûts de développement des applications vocales intégrées au système de chaque assistant sont importants pour les entreprises tierces. En outre, le système reste opaque dans la mesure où les services de plateforme essentiels ont la capacité d'accorder une préférence à leurs propres services en croisant les données entre les services de plateforme essentiels, et notamment les places de marché ou les magasins d'applications, afin de leur procurer un avantage concurrentiel sur les entreprises tierces.

D'une certaine manière, remédier à l'autopréférence par l'application du DMA revient à garantir quantitativement la liberté de choix de l'utilisateur. C'est ce qu'il convient de retenir de l'article 6.3 du texte : « *Le contrôleur d'accès autorise et permet techniquement la désinstallation facile par les utilisateurs finaux de toute application logicielle dans son système d'exploitation, sans préjudice de la possibilité pour ce contrôleur d'accès de restreindre cette désinstallation si elle concerne une application logicielle essentielle au fonctionnement du système d'exploitation ou de l'appareil et qui ne peut techniquement pas être proposée séparément par des tiers.*

Le contrôleur d'accès autorise et permet techniquement la modification facile par les utilisateurs finaux des paramètres par défaut de son système d'exploitation, son assistant virtuel et son navigateur internet qui dirigent ou orientent les utilisateurs finaux vers des produits et des services proposés par le contrôleur d'accès. Pour ce faire, il invite notamment les utilisateurs finaux, au moment de leur première utilisation de son moteur de recherche en ligne, son assistant virtuel ou son navigateur internet énuméré dans la décision de désignation conformément à l'article 3, paragraphe 9, à choisir dans une liste des principaux fournisseurs de services disponibles, le moteur de recherche en ligne, assistant virtuel ou navigateur internet vers lequel le système d'exploitation du contrôleur d'accès dirige ou oriente les utilisateurs par défaut, et le moteur de recherche en ligne vers lequel l'assistant virtuel et le navigateur internet du contrôleur d'accès dirige ou oriente les utilisateurs par défaut ».

Cet article pose les conditions préalables pour aller plus loin dans la démarche en s'attelant aux services proposés en luttant contre les écosystèmes propriétaires, et laisse entrevoir qu'un assistant vocal puisse être livré « vierge » ou possiblement en capacité de permettre la désactivation ou le « déparamétrage » des applications vocales installées par défaut. L'article 6.3 permet également à l'utilisateur de réaliser ses achats en ligne en choisissant son système d'exploitation, son assistant vocal, en sollicitant l'application vocale de son choix sans qu'elle ne lui soit imposée par un mécanisme de classement des contenus relevant des systèmes d'IA et privilégiant les services de plateformes essentiels relevant de l'écosystème de l'assistant vocal (places de marché et magasins d'applications).

²⁴⁹ ARCEP, *Smartphones, tablettes, assistants vocaux... Les terminaux, maillon faible de l'ouverture d'Internet. Rapport sur leurs limites et sur les actions à envisager*, février 2018, p.42.

Enfin, les considérants 62 et 63 venant éclairer le sens de cet article s'intéressent à la capacité des utilisateurs à se désabonner, ce qui permet de rendre effectifs les dispositifs de lutte contre l'autopréférence. En effet, il est précisé « *Afin de préserver la liberté de choix des entreprises utilisatrices et des utilisateurs finaux, un contrôleur d'accès ne devrait pas être autorisé à rendre inutilement difficile ou compliqué, pour les entreprises utilisatrices ou les utilisateurs finaux, le désabonnement d'un service de plateforme essentiel. Il convient de ne pas rendre la clôture d'un compte ou le désabonnement d'un service plus compliqués que l'ouverture de ce compte ou l'abonnement à ce service* ».

La seule application du DMA aux assistants vocaux permettrait, en levant les verrous anti-concurrentiels et techniques, de garantir la liberté de choix de l'utilisateur entre les applications vocales par défaut qu'il n'aura pas désinstallée et celles qu'il aura installées, qui sans classement pré-établis, seront mises à sa disposition sur commande vocale de sa part. Cependant, les obligations posées en vue d'établir la libre concurrence ne font que policer le système de manière préventive sans véritable asymétrie, en dehors de l'article 8 du DMA qui prévoit l'application partielle des obligations en fonction des enjeux à saisir. Elles n'ont pas pour objectif premier d'instaurer la diversité. Seule une véritable approche par la régulation *ex ante* asymétrique permettrait de poser auprès de chaque contrôleur d'accès par effet ricochet à ses services de plateforme essentiels, des obligations ajustées en fonction de l'objectif de diversité.

Mais cette liberté de choix reste conditionnée : elle dépend de briques techniques, de la désinformation, et de dispositifs qui peinent à garantir le pluralisme culturel.

Liberté de choix et accès aux terminaux. Le règlement européen sur l'internet ouvert²⁵⁰ considère Internet, comme « *une plateforme ouverte d'innovation facile d'accès pour les utilisateurs finals, les fournisseurs de contenus, d'applications et de services et les fournisseurs de services d'accès à l'internet* ». Le Conseil constitutionnel dans le cadre de la décision dite HADOPI²⁵¹ a pu considérer, sur le fondement de l'article 11 DDHC, que la participation à la vie démocratique et l'expression des idées et des opinions, « *implique la liberté d'accéder à ces services* ». Sur ce fondement, le règlement accorde des droits aux utilisateurs, à savoir le droit d'accéder et de diffuser des informations et des contenus en ligne en accédant aux terminaux de leur choix. Cependant, l'internet ouvert y est envisagé seulement sous l'angle de la neutralité des réseaux des fournisseurs d'accès à internet (pratique de gestion de trafic, transparence) et ne vise que l'interdiction de restreindre l'utilisation de terminaux connectés aux réseaux.

²⁵⁰ Règlement (UE) 2015/2120 du Parlement européen et du Conseil du 25 novembre 2015 établissant des mesures relatives à l'accès à un internet ouvert.

²⁵¹ Décision n° 2009-580 DC du 10 juin 2009, Loi favorisant la diffusion et la protection de la création sur Internet.

Or l'ARCEP avait, dès février 2018 et à la suite d'une enquête réalisée entre 2015 et 2018, mis en perspective que les terminaux jouent un rôle significatif sur le marché en proposant des contenus, dont les suites à donner se retrouvent dans les dispositions du DMA. C'est le cas notamment des assistants vocaux, clairement visés par l'ARCEP, mais aussi des éditeurs de systèmes d'exploitation, de navigateurs et de magasins d'applications. L'accès au réseau de l'utilisateur à l'aide de l'assistant vocal se trouve en effet conditionné par l'existence de nombreuses briques logicielles, dont l'enjeu pour les acteurs professionnels concernés procède de l'échange de données. La conséquence première est que cet utilisateur n'est pas en mesure d'apprécier toutes les fonctionnalités de son assistant vocal lors de la prise en main. Dès lors, il peut naturellement se laisser imposer les choix des constructeurs sans être en capacité de les analyser. Naturellement cependant, l'absence d'affichage en raison de l'interaction par la voix limite les possibilités d'accéder à une information exhaustive et entraîne une sélection des informations présentées par les entreprises utilisatrices puisqu'il convient de développer des applications vocales. Ces limites résultent de la nature de l'objet mais dénaturent l'accès à Internet en le fragmentant. Sachant par ailleurs que les limites liées à l'interface de l'assistant vocal se répercutent sur l'ensemble des terminaux auquel il permet d'accéder. C'est particulièrement le cas lorsqu'il s'agit d'enceintes intelligentes dans la mesure où tous les contenus qu'elles pourraient diffuser n'ont pas vocation à passer par l'enceinte en l'absence d'écran. On peut considérer que ces enceintes donnent accès à tous les objets connectés grand public en servant de relais. Ou à l'inverse, considérer que l'enceinte vocale limite l'intégration aux seules applications vocales dans le cadre d'un simple système d'interfaçage transparent connu des utilisateurs (ex : Sonos).

L'enjeu des véhicules nouvelle génération. On peut également considérer que le développement d'applications vocales pourrait accompagner de nouveaux usages. Le secteur automobile en constitue un bon exemple : les constructeurs intègrent désormais « *la technologie « voice first » dans leurs nouveaux modèles pour des commandes techniques liées au fonctionnement du véhicule* » et « *embarquent des assistants vocaux pour gérer le divertissement (musique, radio, etc.) et l'organisation de l'agenda (appels, rendez-vous, etc.), soit en les développant en interne, soit en signant des partenariats avec des entreprises technologiques, comme Amazon (Echo Auto), Apple (Car Play) et Google (Android Auto)* »²⁵². Alors que la voiture a toujours été un marché important pour les éditeurs de radios, ces derniers doivent désormais se positionner à l'égard des constructeurs pour apparaître sur les dashboards équipant les véhicules dernière génération. L'enjeu est essentiel dès lors que les véhicules nouvelle génération sont équipés de dashboard permettant d'accéder par la voix à divers services et contenus, particulièrement les services radio et les services d'abonnement musicaux. Dans la mesure où l'usage de la radio est majoritairement mobile, les éditeurs de services de radio doivent être visibles et référencés sur le dashboard du véhicule. Cela les enjoint de réaliser d'importantes négociations

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

commerciales notamment avec les constructeurs de véhicules pour apparaître sur le dashboard. Cette question se retrouve aujourd'hui s'agissant des fabricants de TV connectées au sujet de la place de certains services sur les « boutons » de télécommande de ces téléviseurs. Une bataille s'est engagée au sujet du positionnement des éditeurs de services audiovisuels sur ces interfaces. Ce sont alors les fabricants de ces objets connectés qui jouent le rôle d'intermédiaire entre ces services et l'utilisateur. On retrouve ainsi deux enjeux pour les éditeurs de service : la pluralité des négociations à mener dès lors qu'il existe un grand nombre de fabricants d'objets connectés ; la difficulté de mener ces négociations de façon équilibrée d'autant que ces fabricants opèrent au niveau mondial et pourraient ne pas vouloir adapter chacune de ces interfaces au marché en cause. S'agissant spécifiquement des dashboard des véhicules remplaçant les autoradio traditionnels, il pourrait être envisagé d'imposer la conservation d'un accès hertzien garantissant l'accès au service qui n'apparaît pas sur le « bouton » du dashboard. Une autre solution consisterait à permettre le paramétrage du dashboard par l'utilisateur à partir d'une liste de services (v. *infra*). L'utilisateur informé n'ignore aucun de ces enjeux mais s'oriente naturellement vers un équipement qui répond à ses divers usages d'internet. Le fait est que ce choix, même en l'absence d'action délibérée des professionnels, l'enferme dans un écosystème et dans une habitude d'écoute parcellaire. Or, le choix de l'assistant vocal doit être dissocié de celui des contenus et services utilisés. L'objectif est qu'à terme ils deviennent équivalents en application du principe de non-discrimination.

3. Pluralité des contenus

De l'honnêteté de l'information à la désinformation. Le principe de pluralisme interne qui suppose la diversité ou, à tout le moins, la pluralité des contenus ne peut s'entendre que dans le cadre de la mise à disposition d'une information honnête. Dans le domaine de la communication audiovisuelle, le principe d'honnêteté de l'information suit la préservation du pluralisme. Garanti dans le domaine de la communication audiovisuelle, il prend la valeur d'un impératif constitutionnel²⁵³, sur lequel le CSA et désormais l'ARCOM a pour mission de veiller au même titre que le pluralisme²⁵⁴. L'impératif, à forte teneur déontologique, ne saurait souffrir d'aucune dérogation et s'oppose à des exigences qui n'auraient qu'une portée aléatoire²⁵⁵. Généralement, les conventions rédigées par le CSA et aujourd'hui l'ARCOM sont à ce sujet très précises et imposent le respect de ce principe pour l'ensemble des programmes²⁵⁶. L'honnêteté n'est pas un principe de vérité, mais il suppose que celui qui met à disposition un contenu tende au respect de la vérité à la fois sur le fond et en la forme²⁵⁷. L'honnêteté suppose également de recueillir les différents points de vue même s'ils sont

²⁵³ Cons. const., n° 86-217 DC, préc.

²⁵⁴ Lecture combinée des articles 1 et 28 de la loi n°86-1067 du 30 septembre 1986 préc.

²⁵⁵ V. A. Guedj, *Liberté et responsabilité du journaliste dans l'ordre juridique européen et international*, thèse Paris II, 2000.

²⁵⁶ G. Weigel, « Les dispositions audiovisuelles de la loi 'indépendance des médias' du 14 novembre 2016 étaient-elles nécessaires ? », préc.

²⁵⁷ V. aff. de la « fausse interview de Fidel Castro », CA Paris, 1^{er} ch. A, 24 févr. 1998.

contradictoires, l'objectif étant que le public puisse se forger son opinion et réaliser ses choix de manière éclairée. Cet objectif a disparu de la terminologie utilisée par le législateur, préférant désormais saisir les enjeux de la désinformation. Cependant, le périmètre de la désinformation reste identique, visant à la fois le contenu mais également la manière dont celui-ci est mis à disposition de l'utilisateur, ce qui peut questionner la visibilité des interfaces tout comme les propos illicites ou la question des quotas. D'ailleurs, les processus de désinformation sont parfois imbriqués. Cette approche globale par l'honnêteté, versant positif, ou la désinformation, versant négatif, permet de saisir toutes les facettes de la liberté de choix de l'utilisateur et d'afficher les enjeux contemporains de la société par un choix terminologique.

Liberté de choix et désinformation. Les limites techniques de l'assistant vocal, ayant vocation à disparaître dans le cadre d'une forme plus élaborée d'interface virtuelle, peuvent contribuer à la désinformation de l'utilisateur. Par ailleurs, la fabrication de la réponse et sa restitution par une voix de synthèse sont le résultat d'un long processus opaque que l'on peut saisir sous l'angle des interfaces de choix trompeuses (*dark patterns*), de la publicité ciblée, et tout autre procédé illicite de nature à induire en erreur l'utilisateur dans sa capacité à opérer des choix. L'essor de la désinformation est un autre phénomène lié aux grandes plateformes, qui modèrent en déclassant des contenus selon des dispositifs qui échappent à l'utilisateur.

La lutte contre la désinformation reposerait sur une autre approche du pluralisme, davantage qualitative sur laquelle s'appuie la CEDH. En droit européen la notion de démocratie, bien qu'elle représente un élément fondamental de l'ordre public européen, n'a pas reçu de définition précise compte tenu de la diversité des cultures nationales. La Cour développe une jurisprudence « *plus insistante sur les droits des minorités que sur le pouvoir de la majorité* »²⁵⁸ et contribue à faire entendre toutes les idées même celles qui heurtent et celles qui choquent dans la mesure où le pluralisme exprime la tolérance et l'ouverture d'esprit²⁵⁹. Cette approche est intéressante en ce qu'elle caractérise un enjeu de découvrabilité des contenus, et une réflexion sur leur licéité car le pluralisme ne peut s'entendre au regard des contenus illicites.

Au sens du DSA, le contenu illicite « *couvre les informations relatives aux contenus, produits, services et activités illégaux. En particulier, cette notion devrait être comprise comme se référant à des informations, quelle que soit leur forme, qui, en vertu du droit applicable, sont soit elles-mêmes illicites, comme les discours haineux illégaux ou les contenus à caractère terroriste et les contenus discriminatoires illégaux, soit rendues illicites par les règles applicables en raison du fait qu'elles se rapportent à des activités illégales. Il peut s'agir, par exemple, du partage d'images représentant des abus sexuels commis sur des enfants, du partage illégal d'images privées sans consentement, du harcèlement en ligne, de la vente de produits non conformes ou contrefaits, de la vente de produits ou de la fourniture de services en violation du droit*

²⁵⁸ E. Casenove, *Ordre juridique et démocratie dans la jurisprudence de la CEDH*, thèse Amiens, 1994, p. 161.

²⁵⁹ CEDH, 7 déc. 1976, *Handyside c/Royaume-Uni*, préc.

*en matière de protection des consommateurs, de l'utilisation non autorisée de matériel protégé par le droit d'auteur, de l'offre illégale de services de logement ou de la vente illégale d'animaux vivants. En revanche, la vidéo d'un témoin oculaire d'une infraction pénale potentielle ne devrait pas être considérée comme constituant un contenu illicite simplement parce qu'elle met en scène un acte illégal, lorsque l'enregistrement ou la diffusion au public d'une telle vidéo n'est pas illégal en vertu du droit national ou du droit de l'Union. Il importe peu à cet égard que l'illégalité de l'information ou de l'activité procède du droit de l'Union ou du droit national conforme au droit de l'Union et il est indifférent de connaître la nature ou l'objet précis du droit en question »²⁶⁰. C'est bien sous l'angle de l'accès à une information sûre, prévisible et fiable, qu'il convient de lutter contre la désinformation sous toutes ses formes, et pas sous l'angle des contenus *per se*. L'objectif du DSA n'a pas pour ambition de policer le net et d'éradiquer tous les contenus dérangeants ou les services frauduleux, mais de les limiter et travailler *a priori* sur les comportements afin que de tels contenus disparaissent progressivement par des mécanismes de régulation *ex ante*. Pour autant des choix sont à opérer en fonction des contenus et services visés, et des comportements des intermédiaires. On remarquera à cet égard que les textes récemment adoptés ou encore en discussion, s'agissant des délits de presse voire de la lutte contre les contenus contrefaisants, soulèvent des questionnements convergents (quelle responsabilité des plateformes ? comment notifier le contenu ? quel impact du filtrage algorithmique ? comment articuler les différents droits fondamentaux ?). Certaines de ces interrogations trouvent une réponse *ex post*, mais pour la plupart une identification *ex ante* des procédés illicites peut conduire à distinguer à la fois la réponse à apporter, les moyens à mobiliser ou bien les acteurs impliqués. Or c'est bien toute la difficulté car, en l'état du DSA, le législateur européen se saisit à la fois des plateformes et des contenus/services illicites, sans aller au bout de la logique. La difficulté provient de ce que le dysfonctionnement peut tenir non pas uniquement du fonctionnement des plateformes ou de la circulation de contenus illicites mais bien de l'ensemble du système qu'il convient de mieux comprendre pour mieux saisir. Dès lors, le DSA n'a pas été initialement élaboré pour appréhender l'assistant vocal, ni l'ensemble du système qu'il est censé réguler.*

Pour les assistants vocaux, l'enjeu de désinformation peut être par exemple abordé par la promotion de certains discours dont le caractère illicite peut résulter tout à la fois du contenu (et notamment la promotion de contenus partisans par le gouvernement lors du premier confinement) mais également de la captation frauduleuse de voix connues pour conférer davantage de lisibilité aux propos. Il peut être également abordé sous l'angle de la promotion de certains services lors du choix préalable proposé lors du paramétrage. A ce titre, les réponses sont variées.

Une première réponse globale pourrait résulter, selon les préconisations de l'ARCEP, dans la régulation par la donnée car elle contribue à irriguer le système et il

²⁶⁰ Considérant 12.

conviendrait de disposer des données nécessaires à l'information de tous les intermédiaires techniques pour garantir cette diversité par la mise à disposition de contenus fiables et respectueux des droits fondamentaux (v. *infra*).

Mais, de manière plus symptomatique, il s'agirait de plutôt promouvoir certains services considérés comme intérêt général afin de mettre en visibilité des services et des sources fiables. A ce titre, la directive « Services de médias audiovisuels » (SMA) révisée²⁶¹ propose des règles relatives à la mise en visibilité de contenus qualifiés d'intérêt général. Elle dispose que les États membres sont libres de prendre des mesures pour assurer la visibilité appropriée des services de médias audiovisuels d'intérêt général (art. 7 bis).

Liberté de choix et accès aux contenus culturels. Préserver la liberté de choix de l'utilisateur sous l'angle du pluralisme pourrait également revenir à préserver la diversité de l'offre dans le domaine culturel, par le respect de quotas. S'agissant de la diffusion de contenus audiovisuels, la lecture combinée des articles 33 et 70 de la loi n°86-1067 du 30 septembre 1986 modifiée, précise ainsi que les éditeurs de service s'engagent à diffuser au moins 60% d'œuvres d'expression originale européenne, dont 40% d'œuvres d'expression originale française. A l'échelle européenne, la Directive SMA (UE) 2018/1808 met également en place des quotas. Dès le dispositif initial dénommé *Télévision sans frontières* 89/552/CEE du 3 octobre 1989, la Communauté économique européenne avait institué un système de quotas applicable par les chaînes de télévision et de radio à l'initiative de la France, de manière à contraindre les entreprises de communication audiovisuelle à protéger et constituer un patrimoine européen par la diffusion d'au moins 50% d'œuvres européennes. Il s'agissait à l'époque d'un marché linéaire et globalement maîtrisé par les Etats membres en raison du nombre limité d'acteurs. L'objectif du dispositif procédait de la volonté de construire un véritable marché intérieur de l'audiovisuel en faisant tomber les frontières géographiques. Le changement de logique opéré en 2007 lors de la révision 2007/65/CE du Parlement européen et du Conseil du 11 décembre 2007, en raison du passage à la logique de service entraînée par la délinéarisation de la programmation audiovisuelle, donne naissance aux SMA mais ne remet pas en cause l'intention protectrice des Etats. Lors de la révision de 2018, ce dispositif est complété à l'égard des SMAD. Les fournisseurs de services de médias audiovisuels à la demande doivent veiller à ce que leurs catalogues contiennent une part minimale de 30 % d'œuvres européennes et qu'elles soient suffisamment mises en valeur par des systèmes algorithmiques. Les quotas de musique française à la radio ont permis indéniablement aux auteurs francophones de disposer d'une large exposition. La question des quotas reste manifestement centrale au sein de la réglementation, laquelle n'a d'ailleurs pas vocation à préserver la diversité culturelle mais simplement

²⁶¹ Directive UE 2018/1808 du Parlement européen et du Conseil du 14 novembre 2018 modifiant la directive 2010/13/UE visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à la fourniture de services de médias audiovisuels (SMA).

la concurrence et de manière très variable selon qu'il s'agit d'un Etat membre de l'Union ou d'un Etat tiers.

Les quotas : un mécanisme dépassé ? Les révisions successives du dispositif européen n'ont ni eu pour effet de lutter durablement contre le forum shopping, ni de préserver une culture européenne ou française au sein des écosystèmes. La mission du CSPLA sur les dispositifs de recommandation des œuvres auprès des utilisateurs des plateformes en ligne, conduite par les professeures Valérie-Laure Benabou et Joëlle Farchy²⁶², montre que les quotas ne sont pas d'une grande utilité en matière de musique, et qu'ils sont inadaptés aux services de plateformes dès lors qu'il est aisé par le biais de la recommandation de répondre à la demande au-delà des prescriptions réglementaires. C'est d'ailleurs ce que font Amazon et Netflix, avec l'aide d'utilisateurs allemands ou italiens, particulièrement friands de contenus européens. Le système de la réponse unique, qui dépasse le cadre de la simple recommandation, conforte ce constat. Le rapport met en évidence que la diversité, telle qu'elle ressort de cette réglementation, n'est pas adaptée à l'économie des services à la demande. Nul besoin que les œuvres soient stockées pour répondre à la demande, encore faut-il qu'elles soient découvrables. La présence d'un contenu ne signifie pas qu'il est découvrable et *a fortiori* en raison du mécanisme de la réponse unique. L'assistant vocal n'informe pas l'utilisateur des étapes de son raisonnement ou le périmètre de ses choix, des services activés dans le cadre de la « fabrique » de la réponse. La capacité de filtrage des contenus accentue le risque de ne pas faire émerger des services au sein de l'hyper offre. Le pouvoir de négociation des éditeurs d'assistants vocaux les met en situation de capter une part importante de la valeur et d'imposer les conditions dans lesquelles cette valeur sera restituée. Il existe alors un véritable obstacle à la diversité par déréférencement corrélé à la formalisation de la requête de l'utilisateur. Si ce dernier n'enrichit pas la requête en procédant à une demande simple, la diversité en sera d'autant plus impactée²⁶³. En effet, une requête est dite fermée, lorsque l'utilisateur précise dans sa demande un service spécifique. Dans ce cadre, le référencement est toujours limité par essence. La difficulté résulte alors des requêtes dites « ouvertes », sans précision du service à lancer pour une action précise. L'objet de la requête se limite alors à « *Joue moi telle chanson* » ou « *Lance un flash info* ». Dès lors, le référencement est fondamental car la pratique favorise les marques les plus connues et les acteurs structurants, au détriment d'acteurs émergents ou disposant de capacités financières limitées alors qu'ils proposent des contenus plus diversifiés sans pouvoir justifier de statistiques d'écoutes élevées. C'est à cet endroit que se situe l'enjeu de diversité, en dehors de toute appréciation quantitative qui contribue à asseoir les contenus des acteurs structurants. Vestige de la télévision linéaire, ce dispositif de quotas ne fait pas sens au regard du fonctionnement des assistants vocaux pour lesquels les questions de découvrabilité sont prégnantes.

²⁶² V.-L. Benabou, J. Farchy, *Les dispositifs de recommandation des œuvres audiovisuelles et musicales sur les services en ligne*, rapport présenté au CSPLA, novembre 2021, p. 64&s.

²⁶³ HADOPI / CSA, Etude préc., p.65.

Les critères quantitatifs peuvent-ils être remplacés par des critères qualitatifs ?

Dès lors, les pistes proposées dans le cadre du rapport sur les recommandations mais réfutées immédiatement par les auteurs doivent, à tout le moins, être davantage formalisées. L'appréciation de la diversité procède sans nul doute de choix politiques. Ainsi en est-il du système actuel de quotas qui milite à la faveur du protectionnisme. Cependant, ce système ne peut faire ses preuves qu'appliqué à un territoire circonscrit avec un paysage d'acteurs maîtrisé par les autorités de régulation. En l'absence de quotas, le protectionnisme pourrait s'imposer par un paramétrage des algorithmes, au même titre que la diversité des langues, la nouveauté, les talents émergents, la présence des minorités et des femmes ou la diversité des budgets comme le soulignent les auteurs du rapport. De même que la notoriété d'une production culturelle qui a longtemps été un critère de choix pour les télévisions, le critère du territoire qui permet de valoriser par exemple les spectacles vivants, ou le critère de l'information publique à considérer qu'elle lutte contre la désinformation, sont autant de paramètres à prendre en considération. A ce titre, il n'appartient pas au législateur de mettre en place un nouveau dispositif mais au régulateur de vérifier que ces contraintes éditoriales sont prises en considération et de manière distincte selon les services. Toute la question est de savoir comment les concilier à la personnalisation et de quelle manière le régulateur disposera des moyens pour mettre en œuvre ses engagements dans le respect des choix des utilisateurs.

Visibilité et diversité culturelle (Ordonnance du 21 décembre 2020 transposant la Directive SMA). Formellement, il s'agirait de contrôler que de tels contenus sont automatiquement pris en compte dans la « fabrique » de la réponse par l'assistant vocal et notamment lors de la première phase « d'activation de services » avant que l'assistant vocal ne délivre sa réponse. Pourrait être imaginé, sur le modèle de l'article 7 de la directive SMA, la création de services d'intérêt généraux. La transposition de la directive par l'ordonnance du 21 décembre 2020 confie au CSA (ARCOM) de nouvelles missions en matière de protection de l'intégrité du signal des services de communication audiovisuelle et de mise en avant des services audiovisuels d'intérêt général sur les nouvelles interfaces d'accès aux contenus audiovisuels. En 2020 et sur ce fondement, l'Union européenne de radiodiffusion (UER) appelait les Etats membres à assurer la visibilité des contenus d'intérêt général sur les plateformes audiovisuelles pour « *garantir le droit des citoyens de recevoir des informations exactes et impartiales, pour contrer la désinformation. Sans visibilité, le contenu local, le contenu destiné aux minorités aux auditoires spécialisés et d'autres genres informatifs, (...) ne sont pas promus par les algorithmes. (...) Le contenu d'intérêt général devrait s'étendre à des éléments de contenu désagrégés* ».

En effet, l'article 10 de l'ordonnance que l'on retrouve à l'article 20-7-I de la loi n°86-1067 du 30 septembre 1986 modifiée prévoit que « *Pour l'application du présent article, on entend par "interface utilisateur" tout dispositif présentant à l'utilisateur un choix parmi plusieurs services de communication audiovisuelle ou parmi des programmes issus de ces services* ». Ce dispositif installé sur un téléviseur ou sur un

équipement destiné à être connecté au téléviseur ; installé sur une enceinte connectée ; mis à disposition par un distributeur de services ou mis à disposition au sein d'un magasin d'applications, n'exclut en aucun cas les assistants vocaux. Il doit permettre la mise à disposition de services d'intérêt général dans les conditions précisées par l'ARCOM qui doit s'assurer de leur visibilité (sauf si l'interface propose exclusivement les services d'un éditeur et de ses filiales). Ces services comprennent les services édités par les organismes du secteur audiovisuel public, voire d'autres éditeurs de services de communication audiovisuelle s'ils sont susceptibles de diffuser de tels contenus. A l'usage, l'adjonction d'éditeurs de services privés de communication audiovisuelle semble souhaitable de manière à garantir leur contribution au « *caractère pluraliste des courants et pensée et d'opinion et à la diversité culturelle* », notamment sur les territoires où seul l'accès aux contenus de service public est garanti.

Aux termes de l'ordonnance, au 1er janvier 2022, les opérateurs qui déterminent les modalités de présentation des services sur les interfaces utilisateurs devaient assurer la visibilité appropriée des services d'intérêt général, dont le nombre d'utilisateurs ou d'unités commercialisées sur le territoire français dépasse un seuil fixé par décret. A notre connaissance, aucun décret n'est venu préciser le nombre d'utilisateurs concernés, ou d'unités commercialisées. La liste de ces services n'est par ailleurs pas publiée. La visibilité doit être appropriée aux capacités de personnalisation de l'interface par l'utilisateur. Par conséquent, on pourrait retrouver ces services selon les cas : sur la page ou l'écran d'accueil, dans les recommandations, dans les services de recherches initiés par l'utilisateur, ou *sur les dispositifs de pilotage à distance des équipements donnant accès aux services de communication audiovisuelle*. L'opérateur qui gère l'interface doit rendre des comptes à l'ARCOM sous peine de sanction après mise en demeure.

Quoi qu'intéressant, ce dispositif n'est pas abouti et notamment en raison de l'absence de décret d'application. Plusieurs difficultés apparaissent quant à sa mise en œuvre : quelle visibilité pourrait être envisagée pour les interfaces des assistants vocaux qui permettent la diffusion de ces contenus ? Quelles applications vocales pourraient être considérées comme des services d'intérêt général ? En réalité, la difficulté à lister les services d'intérêt général constitue probablement le nœud du système. A titre d'exemple, sur les 1000 radios françaises, 700 sont associatives. Ne peut-on pas considérer que toutes les radios associatives sont des services d'intérêt général en raison de leur objet, leur mode de financement ? Auquel cas, cela reviendrait à considérer que tous les services de radio sont potentiellement des services d'intérêt général... Par conséquent, il est difficile d'aller plus loin dans la discussion. Et enfin, l'un des écueils de cette directive est de maintenir la règle du pays d'origine, or l'application de la règle du pays de destination est attendue lorsqu'il s'agit de garantir la liberté de choix de l'utilisateur. Reste que le dispositif ne doit pas être écarté pour autant car il doit être combiné à l'action de l'utilisateur au moment du paramétrage (v. *infra*) ou plus exactement de l'initialisation du service. Il faudra engager la conversation

entre la machine et l'humain pour inviter à la diversité. Il faudra par ailleurs, probablement le croiser à d'autres dispositifs afférents aux données que l'on pourrait qualifier d'intérêt général tels qu'envisagé par le texte adopté du règlement sur la gouvernance européenne des données²⁶⁴.

L'ensemble des règles présentées reposent sur des seuils, des quotas qui, en raison de leur caractère objectif, ne permettent pas de saisir d'abord la réalité, à savoir l'extension des écosystèmes par l'utilisation des assistants vocaux, puis la diversité des contenus par la pleine application de l'exigence de transparence notamment au regard du principe de pluralisme qui se mesure encore par un ordre de grandeur dont le respect incombe à l'émetteur. Ce principe semble aujourd'hui intégrer le raisonnement européen croisant la contestation du marché, l'honnêteté de l'information et la diversité des contenus. La diversité reste à construire par une appréciation qualitative non plus du marché cette fois-ci, mais des contenus/services. L'utilisateur contribue grandement à fabriquer cette diversité. Cette appréciation qualitative ne saurait reposer que sur les seuls acteurs structurants du marché. Dans une approche par la régulation prônée par les textes européens, elle est nécessairement diluée sur l'ensemble des acteurs, utilisateurs compris, et ne repose pas seulement sur un dispositif mais sur un ensemble de dispositifs. L'information et les choix de l'utilisateur doivent contribuer à définir la diversité.

B - L'enjeu de la découvrabilité : deuxième levier pour envisager la diversité autrement

Ce que l'on vient de saisir par le pluralisme à travers la lisibilité des interfaces commande de poursuivre la démarche et de repenser la diversité sous l'angle cette fois-ci de la découvrabilité des contenus. Ce sujet est sensible car il exprime un réel paradoxe : l'offre de contenus n'a jamais été aussi abondante sur l'ensemble des services en ligne, qu'il s'agisse de la reprise du signal de télévision linéaire, de la SVOD ou de la TV de rattrapage²⁶⁵. Dans le domaine de la musique, l'offre excède la demande alors que le marché est laissé à quelques acteurs. Ces contenus sont donc disponibles mais sont-ils accessibles pour l'utilisateur ? Aborder cette question dans l'univers des assistants vocaux suppose de mettre en relief la disponibilité des applications vocales pour l'utilisateur mais aussi de revenir sur le filtre de la réponse unique, qui plus est, vocale. L'enjeu de découvrabilité est alors principalement en lien avec le nombre d'applications vocales disponibles.

²⁶⁴ Règlement du Parlement européen et du Conseil portant sur la gouvernance européenne des données et modifiant le Règlement (UE) 2018/1724 (Règlement sur la gouvernance des données), 2020/0340 (COD), adopté en mai 2022.

²⁶⁵ Etude conjointe menée par l'HADOPI et le CSA, *La multiplication des services de vidéo à la demande par abonnement : stratégies de développement et impact sur les usages*, publiée le 9 mars 2021.

La disponibilité d'un contenu pour l'utilisateur n'implique pas sa découvrabilité.

Pour autant, l'exposition des œuvres en dépend²⁶⁶. Une définition a été proposée dans le cadre d'une Mission franco-canadienne : « *La découvrabilité d'un contenu dans l'environnement numérique se réfère à sa disponibilité en ligne et à sa capacité à être repéré parmi un vaste ensemble d'autres contenus, en particulier par une personne qui n'en faisait pas précisément la recherche* »²⁶⁷. Indépendamment de points importants qui peuvent limiter la découvrabilité des contenus mais qui ne sont pas propres aux assistants vocaux (comme la qualité des métadonnées associées aux contenus, les stratégies marketing et de design, les recommandations²⁶⁸, la connaissance des usages par les professionnels ou les politiques publiques de promotion de tels ou tels types de contenus), d'autres éléments spécifiques contribuent à enfermer l'utilisateur dans des choix contraints. Il en va ainsi du nombre d'applications vocales, du caractère vocal de la réponse et de son unicité. Les effets en sont particulièrement amplifiés dans la réponse unique adressée à l'utilisateur. Il est alors loisible d'adapter cette définition aux assistants vocaux, en précisant « *La découvrabilité d'un contenu dans l'environnement numérique se réfère à sa disponibilité en ligne et à sa capacité à être repéré parmi un vaste ensemble d'autres contenus, en particulier par une personne qui n'en formulerait pas précisément la requête ou la question* ».

Le risque d'enfermement. Les conclusions adressées par Joëlle Farchy et Valérie-Laure Benabou dans le cadre du rapport sur les mécanismes de recommandation²⁶⁹ interrogent l'efficacité que tels mécanismes. Le risque avéré reste bien celui de l'autopréférence, lié à celui du classement des contenus, l'enfermement par la recommandation n'y étant pas démontrée. A supposer qu'il existe une « chambre d'écho » dans le domaine de l'accès à l'information, alimentée par des biais de sélection et des mécanismes d'auto-renforcement des convictions de l'utilisateur, elle ne fait que conforter ses tendances naturelles, valorisant des informations qui correspondent à ses convictions. Cependant, la recommandation peut avoir un effet amplificateur d'autant plus important que les éléments de réponse sont limités à une seule réponse. Pour les services culturels, l'intérêt économique passe par la satisfaction immédiate des besoins des utilisateurs, à telle enseigne que les algorithmes de recommandation misent sur la pertinence des résultats et sont peu performants sur le long terme au risque de lasser l'utilisateur qui demande à être surpris. La lassitude est encore amplifiée du fait de la réponse reçue par l'utilisateur à l'issue de son profilage.

Cela pourrait le conduire à changer d'application, à suggérer d'autres contenus en formulant une nouvelle requête. Il faut néanmoins que la requête soit précise. Car tout

²⁶⁶ Rapport établi par S. Clément-Cuzin, inspectrice générale des affaires culturelles et F. Hurard, Inspecteur Général des Affaires culturelles, *La captation de spectacles vivants*, juin 2022, Rapport IGAC, n°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*, rapport présenté au CSPLA, novembre 2021, p. 62 &s.

va dépendre de la commande vocale, de sa formulation à partir de laquelle l'assistant vocal comprend la commande et identifie les réponses disponibles ou va exprimer sa non compréhension de la commande. Selon la technologie utilisée et l'interface utilisateur, soit l'assistant vocal identifie l'option appropriée (la réponse) ou la liste d'options (la liste de réponses possibles). Néanmoins, la Commission européenne dans le cadre de son enquête sectorielle sur l'IoT grand public constate que pour accéder à un fournisseur de services autre que celui par défaut via un assistant vocal, un utilisateur devra généralement ajouter un nom d'invocation spécifique à ses commandes vocales (par exemple "écouter la chanson X sur le service Y"). Ce constat est corroboré par les répondants à l'enquête : « *Nous avons appris que les utilisateurs de commandes vocales utilisent des commandes très génériques et incluent rarement des marques dans leurs commandes. Dans notre cas, les recherches 'génériques' de chaînes musicales sont dirigées vers l'application 'préférentielle' [...]. Ce n'est que lorsque les consommateurs posent des questions très spécifiques pour nos services qu'ils sont dirigés vers nous, ce qui rend la 'découvrabilité' de nos services très difficile* »²⁷⁰.

Le nombre limité d'applications vocales. Cependant, s'agissant de la réponse fournie par l'assistant vocal, la disponibilité des contenus est tout autant questionnée que leur visibilité. La disponibilité est d'emblée liée à des considérations ergonomiques de l'appareil, ou de sécurité limitant l'accès à l'information seulement par la voix²⁷¹. En outre, l'offre de services proposés est encore insuffisante ce qui contraint l'utilisateur. L'étude HADOPI-CSA consacrée aux enceintes connectées précise que « *le choix de la marque de l'enceinte connectée a un impact non négligeable sur les abonnements aux offres culturelles qui seront par la suite associés à ce terminal (...). Au final, 47 % des abonnés musique (soit 15 % des utilisateurs réguliers) et 50 % des abonnés VàDA utilisateurs d'enceintes connectées (soit 21 % des utilisateurs réguliers) ont donc été influencés dans leur choix de service suite à l'adoption de ces appareils, que ce soit positivement (pour s'abonner) ou négativement (pour se désabonner)* »²⁷². Par ailleurs, l'utilisateur de l'enceinte connectée se retrouve enfermé dans son choix initial par un accès limité à des services « *souvent propres à un constructeur. Les éditeurs de contenus devraient pouvoir accéder aux environnements des enceintes connectées dans des conditions équitables et loyales* »²⁷³.

Les assistants vocaux sont intégrés verticalement mais, dans la majorité des cas, il n'y a pas de stratégie d'exclusivité de leur part dans le cadre de l'utilisation d'applications vocales dès lors que les utilisateurs satisfont à leur propre processus de configuration et d'intégration. Il n'y a que Siri qui repose sur un écosystème fermé limitant le choix des applications vocales aux seuls services connexes d'Apple. Ce

²⁷⁰ Commission européenne, Enquête sectorielle préc., point 459.

²⁷¹ ARCEP, Rapport préc., p.36.

²⁷² HADOPI / CSA, Etude préc., p.74.

²⁷³ HADOPI / CSA, Etude préc., p.75.

modèle économique est d'ailleurs remis en cause par les dispositions du DMA. L'absence d'exclusivité est déjà en soi une stratégie pour éviter qu'un nouvel entrant sur le marché des assistants vocaux parmi les acteurs structurants, utilise le modèle de l'exclusivité pour attirer les utilisateurs, ce qui a été pratiqué dans le domaine du jeu vidéo entre Sony et Microsoft²⁷⁴. Alexa permet de lier la clé Firestick d'Amazon et l'enceinte Amazon Echo donnant ainsi accès à tous les services Amazon. Ce constat est prégnant dans le domaine audiovisuel. Mais l'utilisateur peut faire le choix de YouTube et laisser de côté les services proposés par Amazon Prime. Il en va de même pour l'enceinte Google Home comme faisant office de télécommande et permet de piloter son téléviseur par la voix lorsqu'il est jumelé à la clé Chromecast branchée sur le téléviseur ou directement intégrée dans le téléviseur. La synchronisation entre les deux terminaux se fait par le wifi, puis via l'application Google Home téléchargée sur le smartphone, grâce à laquelle on peut ajouter des équipements périphériques et associer des comptes Google²⁷⁵. Rien n'interdit cependant d'accéder aux services proposés par Amazon Prime. La stratégie adoptée sur ce marché est bien celle de la « colonisation » des écosystèmes qui permet, via l'assistant vocal ou l'enceinte connectée, d'étendre l'écosystème initial voir de l'étendre aux autres écosystèmes.

La principale difficulté de la découvrabilité réside dans la fonction d'intermédiation qui pose des obstacles techniques de nature à renforcer la relation de l'utilisateur avec l'assistant vocal au détriment des services proposés. L'utilisateur doit configurer et intégrer les spécificités techniques des services qu'il souhaite utiliser sur l'assistant vocal afin de pouvoir l'utiliser, ce qui suppose que les entreprises tierces aient garanti l'interopérabilité de leur service avec le système technique de l'assistant vocal. Cela fait obstacle au contrôle par les entreprises utilisatrices de l'expérience d'embarquement de l'utilisateur soulignée par la Commission européenne²⁷⁶. En réalité, l'utilisateur va accéder pour la première fois aux services tiers via l'interface de l'assistant vocal, ce qui ne permet pas aux entreprises tierces d'accéder aux données et métadonnées des utilisateurs. D'ailleurs dans la majorité des cas, l'utilisateur dispose d'un compte et d'un identifiant propre à l'assistant vocal ce qui crée une barrière supplémentaire entre l'utilisateur et les entreprises utilisatrices.

La Commission européenne constate encore que la pré-installation des applications constitue un désavantage concurrentiel pour les fournisseurs de contenus dont les applications ne sont pas préinstallées, car un utilisateur doit effectuer des démarches supplémentaires pour accéder à un service qui n'est pas inclus dans l'offre auquel il a souscrit. À l'inverse, les services préinstallés bénéficient de niveaux plus élevés de découvrabilité par les utilisateurs en raison de leur visibilité²⁷⁷. Cette découvrabilité est couplée à la diversité lors ces applications pré installées permettent d'appeler des flux musicaux linéaires, aléatoires ou préconstitués sous forme de playlist. Cela constitue

²⁷⁴ Autorité de la concurrence, Décision n° 15-D-18 du 1er décembre 2015 relative à des pratiques mises en œuvre dans le secteur des jeux vidéo.

²⁷⁵ HADOPI-CSA, Etude préc., p.75.

²⁷⁶ Idem, point 480.

²⁷⁷ Idem, point 449.

à la fois un enjeu et un risque. Multiplier la pré installation d'applications, enferme l'utilisateur dans l'illusion du choix. Cependant, la pré-installation de flux linéaires et pas seulement dans le domaine musical ou la valorisation des agrégateurs de podcasts, méritent réflexion, à condition que l'utilisateur puisse facilement désinstaller ces services selon ses choix. Dans ce cadre contraint, il convient de concilier à l'aide d'algorithmes, la personnalisation des résultats avec la possibilité d'amener l'utilisateur vers d'autres contenus. La personnalisation des choix ne fait pas obstacle à la diversité, tout comme le paramétrage des algorithmes de recommandation qui peuvent intégrer toutes les variables humainement envisageables (augmentation et ajout de contenus recommandés, ajouts de contenus aléatoires, mise en place de systèmes de recommandations hybride)²⁷⁸.

Le classement des contenus est d'autant plus important que la réponse est unique. L'idée de promouvoir la diversité des contenus par la découvrabilité suppose, selon les préconisations de l'ARCEP²⁷⁹, qu'au préalable les critères de référencement et de classement employés par les magasins d'applications vocales connexes aux assistants vocaux soient connus des développeurs. Les magasins d'applications font preuve d'opacité sur les règles de validation et de référencement de ces applications. Dès lors, les développeurs n'ont pas de visibilité sur la mise à disposition de leurs applications. Il conviendrait de garantir une information équilibrée de manière à remédier aux déséquilibres contractuels soulignés par la Commission²⁸⁰. Elle constate que la découvrabilité des applications vocales est contrôlée par les principaux fournisseurs d'assistants vocaux polyvalents, qui ne divulguent pas ces règles aux tiers. Parmi les remontées, un fournisseur de services de contenu créatif explique que *"les fournisseurs de services intégrés ont l'avantage à court terme d'être facilement récupérables via l'intégration sur l'assistant vocal. Cependant, à long terme, cette intégration rend ces services dépendants du pouvoir d'intermédiation de l'assistant vocal. De plus, cette intermédiation se fait sans aucune transparence algorithmique, ce qui finit par obliger les fournisseurs de services à payer s'ils veulent que leur contenu soit visible par les consommateurs"*²⁸¹.

La protection de l'utilisateur contre l'installation d'applications malveillantes. Le DMA remet en cause ce mode d'accès restreint aux applications vocales et notamment le système d'intégration (v. *supra*) pour l'ouvrir à des entreprises utilisatrices tierces. Cependant, un point d'attention concerne les enjeux liés à la protection de l'utilisateur notamment lors de l'accès à des applications malveillantes. Le système d'intégration technique qui met en jeu l'interopérabilité permet notamment de protéger l'utilisateur en ce qu'elles pourraient protéger sa sécurité, la sécurité de son assistant vocal en exploitant les vulnérabilités logicielles ou matérielles, ses

²⁷⁸ Sur le sujet, v. les développements nourris de V.-L. Benabou et J. Farchy, Rapport préc. ; 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, Rapport préc., p. 55.

²⁸⁰ Commission européenne, Enquête sectorielle préc., points 482 et 483.

²⁸¹ Commission européenne, Enquête sectorielle préc., point 481.

données personnelles, voire le prémunir d'une contrefaçon. Par conséquent, soit l'application est malveillante pour l'assistant vocal, soit elle est malveillante pour l'utilisateur. Les applications sont, dans la plupart des cas, paramétrées à l'initiative de l'utilisateur en raison d'une interface qui peut être trompeuse, d'un piratage de l'éditeur de l'application ou par l'utilisation d'une brique logicielle malveillante. En fonction des choix paramétrés par l'utilisateur, différents services peuvent être accessibles par un assistant vocal sans qu'il y ait systématiquement une possibilité d'authentification. Le Pôle d'expertise sur la régulation numérique (PeREN) fournit un éclairage sur les enjeux de sécurité pour la distribution des applications hors des magasins des systèmes d'exploitation²⁸² et détaille les protections qui existent contre ses applications. L'éclairage ne concerne pas spécifiquement les applications vocales, mais le raisonnement est transposable. Il convient de protéger le système d'exploitation de ses vulnérabilités en mettant en place un système de permissions pour tracer le comportement des applications et limiter les fonctionnalités de l'interface afin d'éviter les attaques. De ce point de vue, l'interface vocale épurée de l'assistant vocal pourrait présenter des garanties, mais rien n'est moins certain car elle pourrait se transformer en boîte noire. Dès lors, il convient également de certifier les applications quel que soit leur mode d'installation.

La Commission dans le cadre de l'enquête sectorielle sur les objets connectés grand public précise cependant qu'avant « *d'être disponibles sur un assistant vocal, les applications vocales tierces doivent d'abord accepter les conditions générales de l'assistant vocal et obtenir l'approbation du fournisseur de l'assistant vocal par le biais de son processus d'examen exclusif* »²⁸³. Cette procédure d'intégration/certification n'est pas jugée pertinente par les entreprises utilisatrices tierces de taille modeste. De manière générale, le procédé est lourd et largement unilatéral. La procédure de certification doit associer les développeurs des entreprises tierces, quand bien même l'intégration permettrait de lutter activement contre les applications malveillantes. Mais le système est par trop centralisé.

Catégories de processus de certification pour l'intégration avec les assistants vocaux

a - Les programmes "Works with" ou les certifications d'appareils connectés permettent aux fabricants d'appareils intelligents de rendre leurs appareils domestiques intelligents contrôlables par le biais d'un assistant vocal incorporé dans un autre appareil (c'est-à-dire un haut-parleur intelligent) ou support (c'est-à-dire une application pour smartphone). Pour les fabricants d'appareils domestiques intelligents, ces programmes impliquent généralement l'examen de la documentation du fournisseur d'assistant vocal sur la fonctionnalité de l'API pour l'interopérabilité des appareils domestiques intelligents et le développement d'une application vocale suivant les exigences du fournisseur d'assistant vocal. Le fournisseur d'assistants vocaux teste ensuite l'intégration développée pour garantir une bonne expérience

²⁸² PErEN, février 2022, n°2, https://www.peren.gouv.fr/rapports/2022-02-18%20-%20Eclairage-sur-ouverture-os_FR.pdf

²⁸³ PErEN, préc., point 183.

utilisateur. Si la certification est accordée, les fabricants peuvent utiliser les logos ou les badges "fonctionne avec" sur leur emballage ou pour le marketing en ligne. Exemples de ces processus de certification sont les programmes "Works with Alexa", "Works with Google Assistant" et "Works with Apple Home kit".

b - Les processus de certification intégrée permettent aux fabricants d'appareils intelligents tiers de prendre en charge les assistants vocaux fondés sur le cloud d'Amazon ou de Google sur leurs appareils et d'obtenir les badges "Alexa intégré" ou "Google Assistant intégré". Apple, au contraire, ne propose pas de certifications intégrées, car Siri ne peut être intégré que sur les propres appareils d'Apple. Grâce à l'intégration intégrée, les utilisateurs peuvent accéder à toutes les fonctionnalités prises en charge par l'assistant vocal tiers à partir de l'appareil de maison intelligente du fabricant, y compris le contrôle des appareils intelligents et l'accès aux services tiers par le biais d'applications vocales. À cet égard, les solutions intégrées sont différentes des programmes de travail avec, qui sont limités aux fonctionnalités de contrôle des appareils intelligents. Les processus de certification des solutions intégrées impliquent un test approfondi des appareils intelligents dans lesquels l'assistant vocal sera incorporé par le fournisseur de l'assistant vocal. Sur le plan technique, les fabricants doivent intégrer un microphone intégré et se conformer aux exigences matérielles spécifiées par les fournisseurs d'assistants vocaux.

c - Développement d'applications vocales, telles que les "Alexa skills", les "Google actions", les "Siri shortcuts" ou les "Capsules" pour Bixby, qui fonctionnent sur l'assistant vocal pour lequel elles ont été développées. Les applications vocales rendent les services IoT grand public accessibles à l'utilisateur via les assistants vocaux. D'un point de vue technique, les fournisseurs d'assistants vocaux mettent à disposition des SDK pour construire des applications vocales personnalisées. Les développeurs doivent suivre les directives de l'assistant vocal sur la façon d'écrire une application et se conformer aux règles de compréhension du langage naturel (NLU) de l'assistant vocal. L'application vocale est ensuite soumise à une certification avant d'être publiée et mise à disposition par l'assistant vocal ».

Source : Commission européenne, Rapport final de l'enquête sectorielle sur l'internet des objets pour les consommateurs, 20 janvier 2022, point 204

Le PeREN propose également un système d'audit des applications pour analyser leur comportement avant leur installation²⁸⁴. Il s'agit d'instaurer un mécanisme de tiers de confiance à destination de l'utilisateur. Ce sont les magasins d'applications qui audient les applications qu'ils distribuent mais cet audit peut être révoqué par l'auditeur lui-même ou le développeur de l'application. Ce travail d'audit pourrait tout à fait être effectué directement par des magasins d'applications tiers, ou des organismes d'audits indépendants de la distribution à condition d'en assurer le suivi. La sécurité des assistants vocaux légitime à la fois la pré-installation de certaines applications (à condition que l'utilisateur puisse les désinstaller), mais également les mécanismes de certification, à condition que l'environnement technique et celui des acteurs soient plus ouverts. Parallèlement, le rôle de l'utilisateur est de veiller à

²⁸⁴ PeREN, *Eclairage sur... Applications mobiles : quels enjeux de sécurité pour leur distribution hors des magasins des OS ?* février 2022, n°2, p. 4.

déployer des assistants vocaux sur des équipements à jour et sécurisés de manière à limiter toute malveillance, qui pourrait constituer une brèche importante dans la protection des données personnelles, comme l'identité bancaire.

La voix, seulement la voix. L'utilisation de la voix a un effet structurel d'enfermement incontournable ; l'utilisateur devant conserver à l'esprit que l'interaction se réalise seulement par la voix. Tous les outils concourant à l'information de l'utilisateur pouvant être associés à l'assistant vocal pour la compréhension du traitement des données et l'apprentissage de l'écosystème, altèrent l'interaction vocale en la rendant moins fluide. Pour autant, ces mécanismes portés au titre de la protection des données personnelles pour l'acquisition du consentement, ou au titre de la liberté de choix de l'utilisateur, sont déterminants. Dans le cadre d'un « marché de la voix » en pleine expansion, il reste innovant avec l'arrivée demain de véritables chatbot, dont la puissance des systèmes algorithmiques laisse entrevoir qu'ils seront capables de percevoir la charge mentale et émotionnelle de l'utilisateur à travers la collecte de données dans le cadre de l'interaction vocale. C'est pourquoi il est essentiel de ne pas limiter la question de la diversité à la découvrabilité des contenus mais de concevoir qu'une plus grande transparence serait acquise notamment par une approche de la régulation « par la data » préconisée par l'ARCEP, de manière à ce que le niveau d'information soit comparable pour les utilisateurs, les particuliers et les professionnels²⁸⁵. Si l'information se fait à tous les niveaux, l'utilisateur sera prêt pour passer d'un écosystème à l'autre dans le cadre de la portabilité des données et de l'interopérabilité technique, mais l'on pourra également lever les restrictions liées à l'intégration des assistants vocaux, ou hiérarchiser de manière aléatoire et alternative les contenus et services mis à disposition (v. infra).

Prise en compte de la diversité dans l'offre et dans la demande. Cette démarche est séduisante et particulièrement discutée dans un contexte européen au sein duquel l'approche écosystémique permet de créer une interdépendance entre les marchés, les services, les algorithmes, les données et les acteurs au rang desquels se trouve l'utilisateur. Interroger la diversité sous l'angle de l'offre limite le choix des outils propres à la concevoir. Ainsi, l'envisager dans le cadre de la demande permet de compléter utilement le cadre d'intervention de l'ensemble des acteurs en mettant l'utilisateur en pouvoir d'agir.

C - L'utilisateur : nouvel acteur du pluralisme et de la diversité

Pour qu'il y ait communication entre deux personnes, il faut un émetteur et au moins un récepteur. Cette communication bilatérale suppose que ces deux acteurs, pour être

²⁸⁵ ARCEP, Rapport préc., p. 52&s.

en mesure d'interagir, puissent disposer des mêmes prérogatives. Or, le récepteur a longtemps été considéré comme passif et grandement ignoré du législateur.

Vers un droit à l'action de l'utilisateur. Legs de la Révolution française, le libellé de l'article 11 de la Déclaration des droits de l'homme et du citoyen (DDHC) est remarquable de précisions. Il vise la liberté de l'émetteur, seul sujet de droits et débiteur d'obligations, et non celle du récepteur, considéré comme passif, même si le caractère universel de la DDHC a permis d'étendre la liberté de communication à tous moyens qui procèdent de l'extériorisation de la pensée. Le Conseil constitutionnel considère que l'article 11 est applicable à l'ensemble des communications électroniques²⁸⁶. La qualité de récepteur n'a jamais conféré la défense d'intérêts juridiquement protégés. Le droit de la communication s'est construit dans ce déséquilibre, reposant sur la protection du récepteur par l'émetteur, professionnel de la communication en position dominante. La Convention européenne des droits de l'homme, en son article 10, tout en reconnaissant la liberté de réception et l'accès à l'information, n'augure pas d'une plus grande justiciabilité du récepteur, neutralisée par la référence faite au public, ou au droit du public à l'information. L'intérêt du public demeure au rang des intérêts à préserver, par l'émetteur, mais également par le juge²⁸⁷. En juin 2020, le Conseil interprète une nouvelle fois cet article 11 en reconnaissant, au côté de la liberté d'accéder aux services numériques, la liberté de s'y exprimer, tout particulièrement sur les réseaux sociaux²⁸⁸. La décision censure un dispositif en raison de l'atteinte qu'il fait peser à cette liberté de s'exprimer par le risque de sur-censure. Il considère ainsi, pour la première fois, le rôle actif des utilisateurs sur le fondement de l'article 11 de la DDHC. Ces utilisateurs, à qui le règlement général sur la protection des données (RGPD) et la loi « informatique et libertés » (LIL) avaient reconnu une certaine mise en pouvoir d'agir (*empowerment*) dans le cadre de la protection de leurs données personnelles, se voient reconnaître les mêmes prérogatives en matière de liberté d'expression. Garantir à l'utilisateur la liberté de s'exprimer sur les réseaux sociaux a pour conséquence de subjectiviser son rôle. Par conséquent, la liberté de choix qui lui avait été garantie par le Conseil constitutionnel au XX^{ème} siècle prend désormais tout son sens. Cela permet d'aller au-delà des pistes visant à promouvoir l'enjeu de diversité seulement sous l'angle des plateformes en considérant la diversité de la demande. L'utilisateur dispose donc d'un rôle à jouer dans la caractérisation de la diversité. Il s'agit de lui garantir un droit à l'action pour qu'il soit en capacité de faire ses choix.

Plusieurs niveaux d'action sont nécessaires face à l'effet de réseau et aux effets d'échelle sur l'utilisation massive des données de manière à ce que l'utilisateur puisse être mis en pouvoir d'agir. Il faut tout d'abord envisager sa mise en pouvoir d'agir par la donnée afin que l'utilisateur devienne acteur de sa liberté de choix, ce qui suppose

²⁸⁶ Par ex., Cons. const., 3 mars 2009, n° 2009-577 DC, consid. n° 3, *D.* 2009. 884, point de vue A. Levade, et 2010. 1508, obs. L. Gay ; *AJDA* 2009. 617, tribune P. Wachsmann ; *RFDA* 2009. 580, chron. A. Roblot-Troizier et T. Rambaud ; *Constitutions* 2010. 109, obs. A. Barilari, et 238, obs. M. Disant.

²⁸⁷ Sur ces points, K. Favro, *Droit de la régulation des communications numériques*, LGDJ, coll. Systèmes, 2018.

²⁸⁸ Cons. const., 18 juin 2020, n° 2020-801 DC.

alors une intervention du régulateur (1). Au-delà, il convient de considérer d'autres modalités pour repenser ses interactions avec le fournisseur de l'assistant vocal (2).

1. La mise en pouvoir d'agir de l'utilisateur par la donnée

Régulation par la data. L'ARCEP, dans le cadre de son étude sur les terminaux, a formulé quelques préconisations afin que ces derniers puissent contribuer à l'ouverture d'internet en améliorant la capacité d'agir des utilisateurs notamment dans le choix des terminaux et les possibilités d'en changer²⁸⁹. L'autorité propose de réguler par la donnée afin de garantir non seulement l'information du régulateur pour « *dresser des diagnostics objectifs* », mais également l'information de l'utilisateur, premier pilier de la mise en pouvoir d'agir dans un écosystème biaisé par la capacité des acteurs structurants à capter le marché des assistants vocaux sans garantir effectivement au consommateur final l'accès à l'ensemble des services proposés. « *Une information accrue pourrait permettre de discipliner les comportements des fabricants d'équipements et des éditeurs de systèmes d'exploitation en donnant aux utilisateurs le pouvoir de faire des choix éclairés* »²⁹⁰. L'ARCEP propose de clarifier le champ de l'internet ouvert en posant un principe de liberté de choix des contenus et applications quel que soit le terminal. Centre névralgique de l'écosystème, la data contient en germe toutes les informations utiles aux acteurs du marché, utilisateurs compris, contribuant ainsi à la transparence, la fluidité et la contestabilité du marché. En effet, pour ouvrir les écosystèmes et leurs services de plateforme essentiels, deux actions doivent être menées de concert, conditionnant la mise en œuvre d'outils concrets de régulation *ex ante* dédiés à la mise en pouvoir d'agir de l'utilisateur.

La première action procède du régulateur et tient à sa capacité à collecter l'information auprès des fabricants de terminaux et éditeurs de systèmes d'exploitation. Les informations et données disponibles sont interrogées dans la mesure où elles font défaut à tous les acteurs et pas seulement au régulateur. Les éditeurs d'applications vocales ont peu de visibilité sur les effets de la disponibilité de ces applications sur les systèmes d'exploitation, ce qui comprend notamment les données d'usage et les données de performance du système. Si l'ARCEP préconise une publication de ces données, elle conditionne également la recherche académique et indépendante dans ce domaine en ce qu'elle peut venir en renfort de l'action du régulateur. Or, l'article 40 du DSA prévoit en ce sens un contrôle des très grandes plateformes en ligne et des très grands moteurs de recherche en ligne par des « *chercheurs agréés* » affiliés à un organisme de recherche, cependant, le DMA ne dispose d'aucun dispositif à cet endroit. Même si les deux textes ont un impact considérable sur la gestion des données par leur approche transversale, l'identification, la compréhension et l'évaluation des risques systémiques (justifiant

²⁸⁹ ARCEP, Rapport préc., p.53.

²⁹⁰ Ibid.

l'accès des chercheurs aux données des services et produits proposés par les acteurs structurants) reste l'enjeu fondamental du DSA (article 34 et 35 du texte). Pour autant, une telle démarche eût été légitime pour soutenir la contestabilité des marchés. Une réflexion sur l'expertise du régulateur doit alors être menée afin d'envisager une régulation plus fine des différents écosystèmes et plus particulièrement un accès aux données collectées et traitées par les assistants vocaux.

L'action du PeREN, service à compétence nationale, doit être systématisée et gagner en visibilité à cet endroit. Il constitue, de par les missions qui lui sont attribuées, un centre d'expertise en science des données saisi par les services de l'Etat et les autorités de régulation qui en ont l'utilité (ARCEP, ARCOM, CNIL, Autorité de la concurrence, Ministère de la culture, Ministère de l'économie, des finances, et de la souveraineté industrielle et numérique). Il est amené à intervenir à la fois comme un service d'appui dans le cadre de conventions qui intègrent les particularités réglementaires des autorités administratives visées sur la capacité de collecte des données, mais il intervient également en qualité d'expert dans le cadre de travaux de recherche « *commandités par des services de l'Etat, en réalisant des études à caractère exploratoire ou scientifique : le PeREN agit alors comme un pôle de recherche* »²⁹¹.

Le PeREN a analysé notamment en 2021, dans le cadre d'un partenariat avec la HADOPI, l'offre et la consommation des vidéos proposées sur les chaînes YouTube²⁹² ce qui permet de disposer d'éléments d'analyse suffisamment fins pour caractériser la diversité des contenus mis à disposition par cet acteur, qui de surcroît sont accessibles via les assistants vocaux. De la même manière, le PeREN a fourni un éclairage sur l'AMP (*Accelerated Mobile Pages*), introduite par Google en 2015, désignant un ensemble d'outils visant à accélérer l'affichage de pages web sur mobile de nature à éviter les bloqueurs de publicité, qui ont pu être utiles aux éditeurs de presse²⁹³ ou encore, en 2021, un éclairage sur l'interopérabilité²⁹⁴. Dans ce contexte, le PeREN est en mesure d'analyser un marché et d'en percevoir les évolutions. Son pôle REGALIA est spécialisé par ailleurs dans l'étude des algorithmes. Dans cette perspective, il a délivré notamment une expertise sur la viralité des contenus à la demande de l'ARCOM, sur la minimisation des données obtenues en portabilité pour le compte de la CNIL, sur l'audit des algorithmes en transparence faible, ou sur un dispositif expérimental permettant d'automatiser les analyses pour les services qui ne présentent leurs informations qu'au travers de leurs applications sur smartphone. Ces expertises sont, pour certaines, confidentielles et en tous les cas, initiées par les autorités administratives, et n'ont pas la visibilité ou la spontanéité des travaux académiques de chercheurs. Or, l'expertise du régulateur doit se nourrir de données fines et transversales afin de disposer du bon niveau d'information et être en capacité

²⁹¹ <https://www.peren.gouv.fr>

²⁹² https://hadopi.fr/sites/default/files/2021-12/2021_12_15_rapport_YouTube_PeREN.pdf

²⁹³ PeREN, *Éclairage sur...n°5 – AMP de Google : du dopage au sevrage ?*, septembre 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>

d'intervenir *ex ante* et de manière asymétrique sur le marché. Les conclusions de l'ARCEP dans le cadre de son étude sur les terminaux invitent à cette collecte de données « *pour stimuler l'offre de contenus et services disponibles* »²⁹⁵, et envisagent de recueillir les signalements des utilisateurs, à l'instar de ce qui a été entrepris en matière de communications électroniques, pour réaliser une cartographie des risques générés par les pratiques des acteurs du marché. Le régulateur pourrait ainsi filtrer les attentes de l'utilisateur sans pour autant proposer de les résoudre, mais essentiellement de les documenter. De la même manière, en agrégeant les informations des acteurs tiers, le régulateur peut se doter d'outils de comparaison, qualitatifs ou de notation, des principaux assistants vocaux, afin d'améliorer la surveillance des pratiques des fabricants et des éditeurs de systèmes d'exploitation²⁹⁶. L'objectif avéré est de permettre l'émergence des dix milles acteurs tiers (PME et starts up) dépendant des acteurs structurants.

Fort de ce positionnement, le régulateur doit se voir reconnaître la possibilité d'intervenir dans le cadre de mécanismes de règlements des différends pour traiter les litiges opposant les acteurs structurants aux entreprises utilisatrices tierces. C'est dans une autre perspective que le Règlement 2019/115 prévoit des modalités de règlement des litiges entre les plateformes et leurs utilisateurs professionnels, qu'il s'agisse de l'instauration de mécanismes extrajudiciaires (système de traitement interne de plainte et médiation) ou encore de la possibilité consacrée au profit de divers acteurs d'engager des procédures judiciaires (organismes d'associations représentatives, organismes publics)²⁹⁷. La procédure préconisée par l'ARCEP relève d'une approche par la régulation, et non d'une approche par les pratiques restrictives de concurrence.

ARCEP, Préconisations

« Un régulateur expert pourrait se voir attribuer la compétence de régler les différends relatifs aux terminaux, y compris leurs systèmes d'exploitation, navigateurs et magasins d'applications, et à l'ouverture d'internet. En particulier, les relations entre certains fournisseurs de contenus et services et certains fabricants de terminaux peuvent être déséquilibrées. Il pourrait donc être pertinent que les fournisseurs de contenus et services disposent d'une voie de recours rapide et efficace en cas de litige portant sur leurs conditions d'accès aux terminaux, que ce soit à travers les systèmes d'exploitation, les navigateurs ou les magasins d'applications. Une procédure de règlement de différend pourrait être instaurée pour arbitrer, au cas par cas, les désaccords entre acteurs professionnels concernant :

- les critères de référencement et de classement utilisés par les magasins d'applications,
- la mise en avant de certains contenus et services par un système d'exploitation,
- l'accès de moteurs de recherche tiers aux magasins d'applications,
- l'accès de magasins d'applications tiers aux systèmes d'exploitation,

²⁹⁵ ARCEP, Rapport préc., p.62.

²⁹⁶ ARCEP, Rapport préc., p.54.

²⁹⁷ Règlement 2019/1150, articles 11 à 14.

- l'accès des applications aux API des systèmes d'exploitation ».

Source : ARCEP, Smartphones, tablettes, assistants vocaux... Les terminaux maillon faible de l'ouverture d'Internet, rapport sur leurs limites et sur les actions à envisager, fév. 2018 p.64.

La seconde action, corrélée à la première, suppose que l'information acquise par le régulateur soit mise au service de l'utilisateur afin qu'il comprenne le fonctionnement de l'assistant vocal « *de bout en bout* » pour être en mesure de réaliser des choix éclairés. Cela concerne essentiellement les pratiques des développeurs, ou les critères constitutifs de l'offre des magasins d'applications. En réalité, il est question schématiquement d'informer l'utilisateur de toutes les pratiques susceptibles de remettre en cause un internet ouvert selon les préconisations de l'ARCEP, mais également de lui proposer des outils dédiés à la mise en pouvoir d'agir. Par cette action, il s'agit de transférer à l'utilisateur la garantie du pluralisme des contenus en complément des mécanismes existants reposant jusqu'alors exclusivement sur les éditeurs de contenus.

2. Liberté de choix et interaction avec le service numérique

S'agissant de ses interactions avec le service numérique, l'exercice de la liberté de choix de l'utilisateur doit se réaliser en deux temps. Tout d'abord, celui-ci doit se voir fournir toutes informations nécessaires à son exercice, ce qui conduit à interdire que son choix ne soit trompé par des interfaces de choix manipulatrices ou *dark patterns* (a). Ensuite, l'utilisateur de l'assistant vocal doit être mis en mesure d'exercer son choix en agissant sur l'infrastructure technique, notamment en paramétrant l'outil selon ses attentes (b). Enfin, il doit être en capacité de migrer vers de nouveaux services (c).

a. La liberté de choix conditionnée par l'information de l'utilisateur

Assurer la compréhension de l'utilisateur quant au fonctionnement de l'assistant vocal et des services associés afin de garantir la diversité par la demande suppose tout d'abord d'interdire de tromper le choix de l'utilisateur en ayant recours à des interfaces de choix trompeuses et manipulatrices ; cela nécessite en outre que l'utilisateur soit informé des modalités de classement des contenus.

Interfaces de choix trompeuses (*dark patterns*). Comme la CNIL a pu le relever, « *la conception des interfaces humains-machines, loin d'être une simple question d'ergonomie, comporte des enjeux cruciaux en termes de capacité d'agir et de*

configuration des possibilités de choix »²⁹⁸. Or, les techniques de design peuvent reposer sur l'exploitation des biais cognitifs opérant d'une manière automatique lorsqu'un individu interagit avec un système informationnel et « *des éléments ou mécanismes peuvent être mis en place dans les interfaces pour influencer les décisions des utilisateurs dans un sens qu'ils ne choisiraient pas nécessairement s'ils étaient en face d'une interface loyale et transparente* »²⁹⁹. Garantir la liberté de choix de l'utilisateur d'un assistant vocal ou d'un chatbot suppose donc de porter une attention particulière aux interfaces de choix. Ce point est essentiel compte tenu de la spécificité des assistants vocaux : d'une part, compte tenu de leur interface particulière (interactions vocales, réponse unique et absence d'écran, lien entre l'assistant vocal et son univers comme décrits *supra*) ; d'autre part, en raison des possibilités offertes au fournisseur de l'assistant vocal de personnaliser et d'optimiser les interactions avec ses utilisateurs par l'analyse algorithmique de l'impact de la conception sur l'utilisateur du service à partir de l'analyse d'une grande masse de données d'usage (métadonnées et données à caractère personnel). A cet égard, la CNIL relève que « *ces nouvelles modalités d'interaction se veulent révolutionnaires par la compréhension spontanée qu'en aurait l'utilisateur et la courbe d'apprentissage progressive qu'elles lui offrent* » voire « *ouvrent la porte à des expériences de plus en plus portées sur les émotions* »³⁰⁰.

Une importante réflexion est désormais engagée afin de sanctionner ou d'interdire le recours aux interfaces de choix trompeuses ou manipulatrices dans la mesure où cela peut avoir des effets préjudiciables pour les consommateurs et la concurrence. Il convient tout particulièrement de renvoyer à quatre documents importants publiés en 2022 par la Federal Trade Commission³⁰¹, l'Autorité de la concurrence et des marchés du Royaume-Uni³⁰², la Commission européenne³⁰³ et l'OCDE³⁰⁴ qui font écho aux études précédemment réalisées notamment en France par la CNIL³⁰⁵, en Norvège par le Forbrukerrådet (Conseil de la consommation)³⁰⁶ et plus récemment aux Pays Bas par la Netherlands Authority for Consumers & Markets³⁰⁷. Ces documents présentent diverses typologies des pratiques problématiques voire illicites, qui distinguent par exemple les structures de choix (ex. modifier la présentation des choix à des fins d'autopréférence), l'information sur le choix (ex. rendre l'information difficile à comprendre ou donner trop d'informations) et la pression sur le choix (ex. utiliser les habitudes de l'utilisateur, la pression du temps ou les messages de confiance). L'OCDE relève que « *les dark patterns les plus fréquents sur les sites web et les*

²⁹⁸ CNIL, *La forme des choix. Données personnelles, design et frictions désirables*, Cahier IP n°6, 2019, p. 7.

²⁹⁹ E. Harry, « Dark patterns : quelle grille de lecture pour les réguler ? », *LINC*, 2 septembre 2019.

³⁰⁰ CNIL, préc., p. 9.

³⁰¹ Federal Trade Commission, *Bringing dark patterns to light*. Staff Report. AN FTC Workshop, septembre 2022.

³⁰² Competition & Markets Authority, *Online Choice architecture. How digital design can harm competition and consumers*. CMA 155, avril 2022.

³⁰³ F. Lupianez-Villanueva & alii, *Behavioral study on unfair commercial practices in the digital environment: dark patterns and manipulative personalisation*. Final report, Commission européenne, DG Just, avril 2022.

³⁰⁴ OCDE, *Dark commercial patterns*, OECD Digital economy papers, octobre 2022, n°336.

³⁰⁵ CNIL, *La forme des choix. Données personnelles, design et frictions désirables*, préc.

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.

applications sont le cadrage (présélectionner des choix par défaut ou leur donner une priorité visuelle, cacher des informations ou masquer des publicités), créer un sentiment d'urgence (par des indications de rareté potentiellement trompeuses), générer une preuve sociale (par des indications de popularité potentiellement trompeuses), forcer l'enregistrement ou la divulgation d'informations, inciter à faire un choix ou rendre difficile l'annulation ou l'opt-out »³⁰⁸.

Taxonomie des principaux « dark patterns commerciaux »

Action forcée

Une action forcée cherche à forcer le consommateur à faire quelque chose afin d'accéder à une fonctionnalité spécifique comme s'inscrire ou être contraint de divulguer plus d'informations personnelles que souhaité.

Manipulation par l'interface

Les manipulations par le design visent à susciter des actions spécifiques du consommateur favorables au commerce en ligne par l'organisation de l'information, par exemple en masquant des informations, pré sélectionnant des options, créant une fausse hiérarchie, utilisant une ambiguïté intentionnelle.

Harcèlement/nagging

Le harcèlement implique des demandes répétées au consommateur de faire quelque chose de favorable à l'entreprise, comme activer les notifications ou le suivi de l'emplacement, et peuvent ainsi exploiter la volonté ou le temps limité du consommateur.

Obstruction

Les dark patterns liés à l'obstruction visent à rendre le déroulement d'une tâche ou d'une interaction plus difficile qu'il n'est nécessaire en soi, dans l'intention de dissuader une action, et peuvent donc exploiter l'inertie du consommateur, sa volonté ou son temps limités.

Sournois/Sneaking

Les dark patterns sournois cherchent à cacher, déguiser ou retarder la divulgation d'informations pertinentes pour la décision du consommateur, notamment en ce qui concerne les coûts, et peuvent exploiter l'attention limitée, le biais par défaut, l'effet d'ancrage ou le sophisme des coûts irrécupérables chez les consommateurs.

Pression sociale

Ces interfaces tentent de déclencher une décision sur la base d'observations du comportement d'autres consommateurs, et peuvent donc exploiter le biais de la pression ou du l'approbation sociale.

Urgence

L'urgence impose une limite temporelle ou quantitative, réelle ou fausse, à une offre afin de pousser le consommateur à effectuer un achat, exploitant ainsi l'heuristique de rareté.

³⁰⁸ OCDE, *Dark commercial patterns*, OECD Digital economy papers, octobre 2022, p.5.

Sanctions des interfaces de choix trompeuses. Ces pratiques peuvent être sanctionnées sur le fondement des pratiques commerciales déloyales au sens de la Directive 2005/29 dite PCD³⁰⁹ et, en France, des articles L. 121-1 & s. du Code de la Consommation. On relèvera à cet égard que les orientations de la directive PCD, dans leur version actualisée en décembre 2021, visent spécifiquement les « interfaces truquées » pour désigner le fait de « *faire référence à une sorte de pratique d'encouragement malveillante, généralement intégrée aux interfaces de création numérique. Les interfaces truquées peuvent être fondées sur les données et personnalisées, ou mises en œuvre de manière plus générale, en utilisant des méthodes heuristiques et des biais comportementaux, tels que des effets par défaut ou des biais de rareté* »³¹⁰. Par exemple, une procédure de désabonnement sera considérée comme une pratique déloyale dès lors qu'elle est trop complexe, ou que sont adressés à l'abonné des messages reposant sur l'affectif (« *Nous sommes tristes de vous voir partir* », « *Voici les avantages que vous allez perdre* »), ou encore en cas de recours à des interférences visuelles l'incitant au moyen de diverses images à conserver son abonnement plutôt qu'à y mettre un terme. Cela explique l'évolution récente de la procédure relative au service *Prime* proposé par la société Amazon. Le BEUC (Bureau européen des Unions de consommateurs) a déposé une plainte contre cette société auprès de la Commission européenne en invoquant l'existence de nombreux obstacles au désabonnement de ce service (menus de navigation complexes, formulations alambiquées, choix déconcertants et incitations douces répétées). Depuis, Amazon s'est engagée à modifier son interface pour simplifier sa procédure de désabonnement ; celle-ci doit désormais s'effectuer seulement en deux clics au moyen d'un bouton « annuler » bien visible et non équivoque, afin qu'il soit aussi simple de se désabonner d'un service que de s'y abonner.

Les interfaces de choix trompeuses peuvent en outre faire l'objet de sanction pour violation du droit des données personnelles, comme l'a admis la CNIL dans différentes décisions récentes dès lors que de telles pratiques sont contraires au respect des principes de transparence, de garantie d'un consentement éclairé et du principe de *privacy by default* en orientant les choix des personnes concernées à leur insu, comme a pu le reconnaître la CNIL dans diverses décisions³¹¹ ou encore, plus récemment, le CEPD dans ses lignes directrices consacrées aux *dark patterns* dans les réseaux sociaux³¹².

³⁰⁹ Directive 2005/29/CE du Parlement européen et du Conseil du 11 mai 2005 relative aux pratiques commerciales déloyales des entreprises vis-à-vis des consommateurs dans le marché intérieur.

³¹⁰ Commission européenne, *Orientations concernant la mise en œuvre/application de la directive 2005/29/CE relatives aux pratiques commerciales déloyales*, 29 déc. 2021, 2021/C 526/1, point 4.2.7

³¹¹ CNIL, Délibération 21 janvier 2019, n° 2019-001 ; Délibération 31 décembre 2021, n° 2021-023 et n° 2021-024 ; Délibération n°2022-020, 10 novembre 2022, Discorde INC - Adde, CNIL, *La forme des choix préc.*, p. 39 & s.

³¹² CEPD 3/2022, *Dark Patterns in social media platform interfaces : How to recognize and avoid it*, 2 mai 2022.

Droit de l'UE et principe d'interdiction des interfaces de choix trompeuses. Les réformes du droit de l'Union européenne envisagent désormais d'aller au-delà en édictant un principe d'interdiction des interfaces de choix trompeuses et manipulatrices. Ainsi, l'article 13 du DMA pose une interdiction explicite de contourner les obligations imposées aux opérateurs visés par le texte - les contrôleurs d'accès (gatekeepers) - par des comportements « *qu'ils soient de nature contractuelle, commerciale, technique ou autre, y compris l'utilisation de techniques comportementales ou la conception d'interfaces* ». S'ajoute à cette « règle anti-contournement », une interdiction de détériorer les conditions ou la qualité du service et de rendre l'exercice de droits des utilisateurs ou leur choix excessivement difficile, y compris en proposant des choix de manière partielle, ou encore en utilisant la structure, la conception, la fonction ou le mode de fonctionnement d'une interface utilisateur ou d'une partie connexe pour perturber leur autonomie, prise de décision ou leur libre choix. Il convient toutefois de relever que ce texte ne pose pas une interdiction générale des interfaces trompeuses mais ne concerne que le champ d'application du DMA. En revanche, un plus grand nombre d'opérateurs de l'IoD seront concernés par cette prohibition lorsque sera adoptée la proposition de règlement sur les données, dit « *Data Act* ». Ce texte prévoit en effet, en son article 6, que « *l'opérateur ne doit pas contraindre, tromper ou manipuler l'utilisateur de quelque manière que ce soit, en nuisant ou en portant atteinte à son autonomie, à sa prise de décision ou à ses choix, y compris au moyen d'une interface numérique* »³¹³.

La proposition de règlement dite « *AI Act* » paraît quant à elle bien moins ambitieuse, en l'état actuel des négociations. En effet, celle-ci prévoit que, parmi les pratiques interdites (les autres étant soumises à diverses strates d'obligations en fonction du risque généré), ne figurent que les utilisations qui interviennent « *à des fins d'influencer de manière subliminale le comportement d'une personne en vue de lui causer ou de causer à un tiers un dommage, ou les systèmes qui exploitent la vulnérabilité d'un groupe de personnes en vue de fausser le comportement de l'une de ces personnes et de causer un dommage* ». Or, les conditions très restrictives posées par ce texte ne manqueront pas d'en limiter considérablement la portée, sauf à ce qu'il évolue d'ici à son adoption³¹⁴. L'enjeu est pourtant majeur, particulièrement dans la perspective du déploiement des chatbot et de l'exploitation possible des données émotionnelles à des fins de manipulation de leur utilisateur³¹⁵. On renverra à ce titre aux analyses du Conseil de l'Europe qui, dans sa Déclaration sur les capacités de manipulation des processus algorithmiques, relevait que « *les outils d'apprentissage automatique (machine learning) actuels sont de plus en plus capables*

³¹³ V. également l'article 25 du DSA qui interdit aux plateformes « *de concevoir, organiser ou exploiter leurs interfaces en ligne de façon à tromper ou manipuler les destinataires de leur service ou de toute autre façon propre à altérer ou à entraver substantiellement [leur] capacité à prendre des décisions libres et éclairées* ».

³¹⁴ Sur ce point, v. S. Vergnolle, « L'illusoire interdiction des pratiques manipulatoires dans le projet de règlement sur l'intelligence artificielle », in *Un droit de l'intelligence artificielle : entre règles sectorielles et régime général*, dir. C. Castets-Renard et J. Eynard, Bruylant, à paraître.

³¹⁵ V. notamment CNPEN, *Agents conversationnels : enjeux d'éthique*, avis n°3, nov. 2021, p. 10 - R. Chatelier, « Captation des émotions : comment vous le direz pourra être retenu contre vous... », *LINC* 4 avril 2018.

non seulement de prédire les choix, mais aussi d'influencer les émotions et les pensées et de modifier le déroulement d'une action, parfois de façon subliminale » ; compte-tenu des « niveaux, très subtils, subconscients et personnalisés de la persuasion algorithmique », ils peuvent avoir des effets significatifs sur « l'autonomie cognitive des citoyens et leur droit à se forger une opinion et à prendre des décisions indépendantes »³¹⁶.

Classement et information de l'utilisateur. S'agissant de la diversité, l'un des enjeux majeurs s'agissant de l'information de l'utilisateur concerne le classement des applications vocales mises à sa disposition et partant des contenus culturels auxquels il peut accéder via l'assistant vocal. Cette situation est d'autant plus problématique que l'utilisateur a rarement accès à plusieurs magasins d'applications, mais se limite à celui proposé par l'assistant vocal. Dans ce cas, il est important qu'il puisse être informé du classement des contenus proposés de manière à ce qu'il adapte sa requête et l'enrichisse des précisions nécessaires pour orienter la réponse en fonction de ses goûts.

Classement, recommandation, préférence. Comme cela a été relevé, le classement peut apparaître comme un moyen de favoriser la mise en avant de certains contenus. Le rapport de la mission pour le CSPLA sur la recommandation s'est intéressé, à ce titre, au classement des contenus comme un préalable aux mécanismes de recommandation. L'article 3. s. du DSA définissant les systèmes de recommandation est éclairé utilement par le considérant 62 du même texte disposant que « *la manière dont les informations sont hiérarchisées et présentées sur l'interface en ligne d'une très grande plateforme afin de faciliter et d'optimiser l'accès aux informations pour les bénéficiaires du service revêt une importance capitale pour les activités de la plateforme. Cela consiste, par exemple, à suggérer, classer et hiérarchiser les informations de manière algorithmique, en les distinguant par le texte ou par d'autres représentations visuelles, ou en organisant de toute autre manière les informations fournies par les bénéficiaires* ». Le classement constitue une étape du mécanisme de recommandation, lequel ne disparaît pas de la « fabrique » de la réponse ». Ce mécanisme de recommandation n'est pas essentiel car la hiérarchisation n'est pas connue de l'utilisateur. Elle se rapporte à une architecture de présentation qui va mettre en avant tel ou tel contenu sans nécessairement les recommander au sens propre du terme. « *En d'autres termes, ce n'est pas parce qu'une plateforme décide de privilégier un mode de présentation de certains contenus qu'il s'agit nécessairement de « recommandation »*³¹⁷. Les contenus mis en avant par les assistants vocaux comme nous avons pu l'analyser, peuvent parfois relever de mécanismes d'autopréférence que l'on ne peut assimiler à des recommandations. En revanche, dans toutes ces hypothèses, le classement préalable des contenus reste le

³¹⁶ Conseil de l'Europe (2019), *Déclaration du Comité des ministres sur les capacités de manipulation des processus algorithmiques* - Adde, N. de Marcellins-Warin, F. Marty, E. Thelisson et T. Warin, « Intelligence artificielle et manipulations des comportements de marché : l'évaluation *ex ante* dans l'arsenal du législateur », *RIDE* 2020/2, t. XXXIV, p. 203, spéc. p. 215 s.

³¹⁷ V.-L. Benabou et J. Farchy, *Les dispositifs de recommandation des œuvres audiovisuelles et musicales sur les services en ligne*, rapport final présenté au CSPLA, 2021, p.16.

dénominateur commun, ce qui peut susciter diverses interrogations lorsque le classement conduit à faire primer les contenus des acteurs structurants.

Informé l'utilisateur du classement algorithmique. De nombreux textes européens s'intéressent à la question de l'information sur le classement des produits et services mais ne ciblent pas nécessairement l'utilisateur. La question du classement des produits et services proposés est cruciale dans les écosystèmes de plateforme à l'égard des plateformes elles-mêmes, des entreprises utilisatrices désireuses de voir émerger leurs produits, et les utilisateurs.

A ce titre, une première définition du classement a été proposée par l'article 2 § 8 du Règlement 2019/1150 dit PToB³¹⁸, comme « *la priorité relative accordée aux biens ou services proposés par le biais de services d'intermédiation en ligne, ou la pertinence reconnue aux résultats de recherche par les moteurs de recherche en ligne, tels qu'ils sont présentés, organisés ou communiqués, respectivement, par les fournisseurs de services d'intermédiation en ligne ou par les fournisseurs de moteurs de recherche en ligne, quels que soient les moyens technologiques utilisés pour une telle présentation, organisation ou communication* ». Cependant, cette définition vise les rapports entre professionnels de manière à protéger les entreprises utilisatrices en préservant le secret des affaires. La prise en compte des intérêts commerciaux ne doit pas conduire au refus de divulguer les principaux paramètres qui déterminent le classement. Il ne s'agit pas de promouvoir un fonctionnement détaillé des mécanismes de classement des fournisseurs concernés, notamment des algorithmes³¹⁹. Le règlement impose aux fournisseurs de services d'intermédiation de décrire dans leurs conditions générales les principaux paramètres qui déterminent le classement mais aussi les raisons justifiant leur importance par rapport aux autres paramètres, dans l'objectif d'informer les entreprises utilisatrices. Dans le cas des moteurs de recherche en ligne, les informations relatives aux principaux paramètres doivent être énoncées dans une formulation claire et compréhensible. L'exigence de transparence, même si elle profite au consommateur, est seulement motivée par l'intérêt des entreprises utilisatrices en proposant « *la description générale des principaux paramètres de classement permette aux entreprises d'apprécier leur degré de pertinence et d'avoir une bonne compréhension du fonctionnement du classement dans le contexte de l'utilisation des services d'intermédiation ou des moteurs de recherche. Il n'est pas en revanche nécessaire qu'elles aient une connaissance précise et détaillée du fonctionnement des mécanismes de classement, et notamment des algorithmes utilisés dans ce cadre* »³²⁰.

Si le règlement profite au consommateur/utilisateur, il n'en est pas la cible alors qu'il s'attend légitimement à ce que le résultat de sa requête lors de l'interaction humain-

³¹⁸ Règlement 2019/1150 du Parlement européen et du Conseil du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne.

³¹⁹ V. considérant 23 de la Directive (UE) 2019/2161 et considérant 27 du Règlement 2019/1150.

³²⁰ G. Loiseau, « Le règlement Platform to Business », CCE, n°7-8, 2020, étude 13.

machine soit naturel, ou à tout le moins relativement neutre. Or, il est possible pour les fournisseurs de contenus d'inclure dans le résultat des publicités payantes ou classer prioritairement certains produits en raison des paiements directs ou indirects reçus des professionnels vendeurs de ces produits. C'est pourquoi l'article 2, point m) de la Directive 2019/2161 de protection des consommateurs³²¹ fournit une définition du classement en prenant en compte l'interaction avec le consommateur : « *la priorité relative accordée aux produits, tels qu'ils sont présentés, organisés ou communiqués par le professionnel, quelle que soit la technologie utilisée pour une telle présentation, organisation ou communication* ». La lecture de cet article doit être éclairée par le considérant 19 de la directive, comme « *résultant [notamment] de l'utilisation du séquençage algorithmique, de mécanismes d'évaluation ou de notation, de la mise en surbrillance, d'autres outils de mise en évidence ou d'une combinaison de ces différents moyens* ». L'un des apports de cette directive résulte de l'article 7 § 4 bis qui établit une prescription d'information spécifique au sujet des principaux paramètres déterminant le classement ; « *Lorsque la possibilité est donnée aux consommateurs de rechercher des produits offerts par différents professionnels ou par des consommateurs à partir d'une requête consistant en un mot-clé, une phrase ou la saisie d'autres données, indépendamment de l'endroit où ces transactions sont finalement conclues, les informations générales mises à disposition dans une section spécifique de l'interface en ligne, qui est directement et aisément accessible à partir de la page sur laquelle les résultats de la requête sont présentés, concernant les principaux paramètres qui déterminent le classement des produits présentés au consommateur en réponse à sa requête de recherche, et l'ordre d'importance de ces paramètres, par opposition à d'autres paramètres, sont réputées substantielles. Le présent paragraphe ne s'applique pas aux fournisseurs de moteurs de recherche en ligne tels que définis à l'article 2, point 6, du règlement (UE) 2019/1150 [...]* ».

L'obligation d'information visée reste par trop circonscrite. Elle suppose déjà une requête du consommateur. Cette prescription d'information s'applique uniquement aux professionnels qui permettent aux consommateurs de rechercher des produits offerts par d'autres professionnels tiers ou par des consommateurs, c'est-à-dire les places de marché en ligne et les outils de comparaison. L'obligation d'information s'applique également dans l'hypothèse où le professionnel permet d'effectuer des recherches sur une interface en ligne par commande vocale, au lieu de taper la requête. Toujours dans ce cas, les informations doivent être mises à disposition pour consultation sur le site internet/l'application du professionnel « *dans une section spécifique de l'interface en ligne* ». Même si cela nuit à l'interaction vocale, l'information doit être consultable sur un site internet ou une application spécifique en ligne, aisément accessible depuis la page sur laquelle figure l'offre ou la commande vocale qui a permis d'y parvenir. Les informations doivent alors être fournies de manière claire et compréhensible et sous une forme adaptée à la technique de communication à distance. Il est également précisé que ces informations doivent se

³²¹ Directive 2019/2161 du Parlement européen et du Conseil du 27 novembre 2019 de protection des consommateurs.

trouver dans une section spécifique de l'interface en ligne qui est directement et aisément accessible à partir de la page sur laquelle les offres sont présentées. En revanche, elle ne s'applique pas aux consommateurs qui opèrent des choix et recherchent leurs propres offres de produits différents de ceux proposés par les plateformes structurantes sans entrer en requête, ni aux moteurs de recherche en ligne, définis par le règlement 2019/115. En réalité, en l'absence de requête de l'utilisateur/consommateur, et en présence d'une interface par défaut, la directive ne s'appliquera pas.

Le niveau d'information dans l'interaction avec l'utilisateur. Parce qu'il s'agit de l'utilisateur, l'information doit rester claire et concise. Pour la même raison, les règles de la directive n'exigent pas d'explication des « *raisons* » de l'importance relative des principaux paramètres de classement, comme le fait le Règlement PToB. Si l'information doit rester générale à l'instar du règlement, en fournissant les principaux paramètres déterminant le classement des produits, elle doit l'informer sur l'importance relative de ces paramètres par rapport aux autres paramètres. Cependant le considérant 23 de la directive précise que la description des paramètres de classement par défaut peut rester générale et ne doit pas être fonction des requêtes individuelles. Les deux textes visés imposent aux plateformes d'informer leurs entreprises utilisatrices au moyen d'informations inscrites dans les conditions générales de la plateforme lors des transactions entre entreprises, ou de s'assurer de la disponibilité de ces informations lors de la phase précontractuelle. Il s'agit donc de satisfaire, avant d'aller plus en avant dans la démarche, à l'exigence de transparence.

Les critères du classement. Ces nouvelles règles en matière de transparence du classement sont similaires dans le cadre des deux textes présentés. Le considérant 22 de la directive en schématise parfaitement l'intention, précisant que « *les paramètres déterminant le classement sont l'ensemble des critères et processus généraux, des signaux spécifiques intégrés dans des algorithmes ou d'autres mécanismes d'ajustement ou de rétrogradation utilisés en relation avec le classement* ». Au titre du Règlement PToB, les services d'intermédiation peuvent communiquer à leurs entreprises utilisatrices des informations plus détaillées de manière à expliquer quels sont les paramètres de classement destinés au consommateur. Pour expliciter ces critères de classement et concourir au respect de l'exigence de transparence, la Commission a publié des lignes directrices³²² précisant l'application du règlement PToB. Ces lignes directrices fournissent d'utiles pistes de réflexion³²³.

L'apport du DMA. Le Règlement PToB et la Directive PCD se contentent de poser à l'endroit des plateformes une obligation de transparence renforcée, le DMA prescrit quant à lui, des obligations plus contraignantes, même si elles ne concernent que les

³²² Commission européenne, *Lignes directrices concernant la transparence en matière de classement, conformément au règlement (UE) 2019/1150 du Parlement européen et du Conseil*, (2020/C 424/01), 8 décembre 2020.

³²³ Communication de la Commission Lignes directrices concernant la transparence en matière de classement, conformément au règlement (UE) 2019/1150 du Parlement européen et du Conseil 2020/C 424/01.

seuls acteurs structurants. L'approche de ces différents textes ne poursuit pas la même logique. Ce qui est au centre des préoccupations du DMA procède de ce double rôle d'intermédiation des services de plateformes essentiels à l'égard des entreprises utilisatrices tierces, mais aussi en préservant directement leurs produits et services³²⁴. Ces acteurs structurants remettent en cause la contestabilité de leurs produits et services réduisant ainsi le choix des utilisateurs. L'article 2 du DMA interroge la question du classement au regard des pratiques restrictives de concurrence et spécifiquement de l'autopréférence, visant expressément les assistants vocaux. Cet article définit alors le classement comme « *la priorité relative accordée aux biens ou services proposés par le biais de services d'intermédiation en ligne, de services de réseaux sociaux en ligne, de services de plateformes de partage de vidéos ou d'assistants virtuels, ou la pertinence reconnue aux résultats de recherche par les moteurs de recherche en ligne, tels qu'ils sont présentés, organisés ou communiqués par les entreprises fournissant des services d'intermédiation en ligne, des services de plateformes de partage de vidéos, des assistants virtuels ou des moteurs de recherche en ligne, indépendamment des moyens technologiques utilisés pour une telle présentation, organisation ou communication et indépendamment du fait qu'un seul résultat soit ou non présenté ou communiqué* ». Il ressort de ce texte que les services de plateformes essentiels ne doivent pas s'octroyer un traitement différencié ou préférentiel en matière de classement « *que ce soit par des moyens juridiques, commerciaux ou techniques* ». Les conditions présidant au classement doivent être équitables, couvrant ainsi toutes les formes de priorité relative, même celles issues de contraintes techniques et notamment sont visées les réponses vocales³²⁵. Le contrôleur d'accès ne doit pas traiter de manière plus favorable, en matière de classement ou selon d'autres paramètres, les services et produits proposés par lui-même ou par tout tiers appartenant à la même entreprise, que les services ou produits similaires d'un tiers, et doit appliquer des conditions transparentes, équitables et non discriminatoires à ce classement des produits et services des tiers. Dès lors, les services de plateforme essentiels doivent appliquer à des produits et services similaires aux leurs et fournis par les entreprises utilisatrices des conditions de classement transparentes, équitables et non discriminatoires. Si la découvrabilité du service en dépend, la simple trouvabilité n'est pas assurée³²⁶. Généralement, les accords contractuels qui président à cette mise en visibilité de certains contenus sont particulièrement décriés car les négociations sont multinationales et les services locaux en sont généralement écartés³²⁷. Dans le cadre de son enquête sectorielle, la Commission européenne préconise des paramètres de parité permettant à chaque entreprise utilisatrice d'associer son propre contenu à celui de l'acteur structurant, davantage visible, en retenant par ailleurs la possibilité d'améliorer la visibilité des contenus avec, par exemple, l'adjonction d'un logo ce qui permet d'assurer une promotion des contenus comparables, ou au moins égale. On relèvera qu'à l'occasion

³²⁴ Considérant 43 du DMA.

³²⁵ V.-L. Benabou et J. Farchy, Rapport préc., p.57.

³²⁶ Commission européenne, Enquête sectorielle préc., point 466.

³²⁷ Idem, point 467 a).

de cette enquête sectorielle et dans la perspective de renforcer cette parité, un répondant a « *suggéré que les principaux fournisseurs de plateformes technologiques IoT grand public pourraient être en mesure d'exploiter les algorithmes applicables à leur propre avantage, par exemple en associant des termes de recherche spécifiques à leur contenu de première main afin qu'il apparaisse plus haut dans le classement* »³²⁸.

L'objectif est de permettre à l'utilisateur, par une information appropriée, de dissocier le référencement et le classement des applications vocales en lui permettant d'interagir avec l'ensemble des magasins d'applications pour opérer des choix. Il s'agit de tenir compte de la manière dont les utilisateurs accèdent à leurs services. S'ils y accèdent par l'intermédiaire d'assistants vocaux, il conviendrait de « *prendre en considération la solution utilisée par les utilisateurs de leur service pour accéder à d'autres informations au sujet du service* »³²⁹. En matière de contenu audiovisuel cependant, lorsque l'assistant vocal embarque d'autres services connectés, la difficulté peut tenir à l'existence de boutons de télécommande qui permettent un accès direct à certains services ou à un assistant vocal. Le classement, dans ce cas, est neutralisé par des considérations purement matérielles assurant une préférence visuelle à tel ou tel service par le jeu des interfaces³³⁰. A l'instar des préconisations de l'ARCEP dans le cadre de son étude sur la neutralité des terminaux, il appartient à l'utilisateur informé de contourner « *les algorithmes d'ordonnement des magasins d'applications et être plus actif dans le paramétrage des critères de sélection des applications disponibles au sein des magasins* »³³¹. Il serait alors en capacité de déjouer le classement des contenus qui ont présidé à la réponse.

b. La liberté de choix conditionnée par la possible action de l'utilisateur sur le système

Au-delà de ces informations, l'utilisateur doit également être mis en mesure d'agir sur l'interface avec laquelle il va interagir. Cela suppose tout particulièrement de s'intéresser aux différentes facettes du paramétrage par défaut qui peut concerner le choix de l'assistant vocal *per se*, celui des applications installées ou encore des applications appelées par l'assistant vocal.

Encadrement du paramétrage par défaut. Le paramétrage par défaut, qu'il s'agisse de la pré-installation ou de valeurs définies par défaut, peut servir l'intérêt des utilisateurs de services numériques en leur évitant de procéder à un grand nombre de choix actifs qui leur imposeraient de mobiliser un temps et des ressources considérables.

³²⁸ Idem, point 467 b).

³²⁹ Commission européenne, Lignes directrices préc., point 119.

³³⁰ Commission européenne, Enquête sectorielle préc., point 469.

³³¹ ARCEP, *Smartphones, tablettes, assistants vocaux... Les terminaux maillon faible de l'ouverture d'Internet, rapport sur leurs limites et sur les actions à envisager*, fév. 2018, point 4.4.2.

Il pourra en revanche être jugé trompeur dans l'hypothèse où il vise à fausser leur comportement. En ce sens, les orientations de la Commission sur la Directive PCD soulignent que cette pratique « *a une incidence considérable sur la décision commerciale du consommateur moyen. Les professionnels pourraient non seulement inciter les consommateurs à réaliser certaines actions, mais aussi en effectuer d'autres à leur place, par exemple en utilisant les cases pré-cochées, notamment pour facturer des services supplémentaires, ce qui est interdit par l'article 22 de la DDC* »³³². Le paramétrage par défaut peut en outre fausser le libre jeu de la concurrence, notamment lorsqu'il permet à un opérateur d'exploiter un pouvoir de marché pour conserver ou acquérir de nouveaux clients et favoriser ses services connexes au détriment de services tiers³³³.

Afin d'assurer la contestabilité et l'équité des marchés numériques, le DMA encadre désormais ces pratiques en imposant des obligations *ex ante* aux contrôleurs d'accès au travers diverses dispositions que l'on peut classer en trois catégories³³⁴ plusieurs d'entre elles prolongeant les recommandations déjà formulées par l'ARCEP dans son rapport sur la neutralité des terminaux³³⁵. Certaines doivent permettre un changement facile des valeurs par défaut par les utilisateurs finaux, à l'image de l'article 6.3 du DMA qui impose d'autoriser et de permettre à l'utilisateur de modifier facilement les paramètres par défaut s'agissant notamment de son assistant virtuel. D'autres tendent à exiger l'utilisation d'écran de choix initial par défaut afin de forcer les utilisateurs finaux à faire des choix actifs, ce qui sera imposé ici encore pour les services d'assistant vocal selon l'article 6.3 du DMA. Enfin, le texte entend promouvoir de manière générale le choix de l'utilisateur final afin de favoriser le changement de fournisseur de service et le multi hébergement : pour ce faire, l'article 6.4 doit permettre l'installation de magasins d'applications ou d'applications tiers pouvant être en outre proposés comme valeur par défaut, sous réserve de ne pas compromettre l'intégrité de l'appareil ou du système ; l'article 6.6 impose que l'utilisateur puisse passer d'un service à l'autre sans restriction technique ou autre forme de restriction, et puisse s'abonner à des services différents ; l'article 6.3 lui permet quant à lui de désinstaller des applications du système dès lors qu'elles ne sont pas essentielles à son fonctionnement.

Principe d'efficacité et paramétrage. Ces mesures doivent être lues à l'aune du principe d'efficacité imposé par l'article 8 du DMA dès lors que ce texte enjoint au contrôleur d'accès d'adopter les mesures propres à assurer le respect des obligations imposées au titre des articles 5 et 6 et, plus généralement pour respecter les objectifs

³³² Commission européenne, *Orientations concernant la mise en œuvre/application de la directive 2005/29/CE relatives aux pratiques commerciales déloyales*, 29 déc. 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, avril 2022, point 1.21 - comp. l'affaire Android : Trib. UE, 14 sept. 2022, aff. T-604/18, Google et Alphabet c. Commission européenne.

³³⁴ Classification proposée par A. Fletcher, *DMA Switching tools and choice screens*, Issue Paper, CERRE, novembre 2022, p.7 &s.

³³⁵ ARCEP, Rapport préc., p. 52&s.

du règlement. Comme cela a pu être relevé³³⁶, ces mesures ne seront efficaces s'agissant du paramétrage par défaut que si les utilisateurs saisissent ces opportunités de choix et dès lors que ce choix ne sera pas trop contraignant pour eux. Or, le choix représente généralement une charge mentale importante pour les utilisateurs alors que le design des interfaces est pensé « sans couture », pour rendre l'utilisation du service la plus facile et la plus fluide possible, ce qui est particulièrement le cas pour les assistants vocaux dont l'objet est *per se* conçu à cette fin. D'autant que l'on sait que l'utilisateur a généralement tendance à valider les choix par défaut ou les choix les plus saillants (biais de saillance) ou choisir les noms qu'ils connaissent, ou conserver un service existant (biais de statu quo). L'architecture de choix sera alors déterminante, ce qui explique que le DMA prohibe les architectures de choix trompeuses en son article 13 (v. *supra*).

Il a en outre été souligné qu'il conviendra de ne pas rendre le choix trop complexe à réaliser, ce qui suppose de concevoir un design adéquat et notamment: de proposer une rubrique de choix bien identifiée et aisément accessible ; de consacrer une section aux paramètres par défaut en proposant une liste de choix possibles, sans qu'il ne soit nécessaire de choisir parmi une liste constituée de telle sorte à décourager l'utilisateur parce que trop longue ou trop complexe (v. *supra* pour les services qui pourraient susciter l'intérêt du récepteur à l'image des services de podcast ou de playlists déjà installées qu'il pourra désinstaller s'ils ne les souhaitent pas les conserver) ; de proposer une modalité de choix rendant l'exercice de celui-ci aisé, notamment de prévoir une commande permettant de « *faire de ce service ma valeur par défaut* »³³⁷. On renverra à cet égard aux réflexions menées par la CNIL s'agissant d'un design des interfaces respectueux de la privacy (design by privacy) qui pourrait constituer un modèle à suivre³³⁸.

Modalités de choix et parallélisme des formes. Une attention particulière devra en outre être apportée aux modalités de choix, à savoir la manière offerte à l'utilisateur d'exercer plus ou moins facilement celui-ci. Cela revient à imposer le respect d'un principe de parallélisme des formes entre les services du fournisseur d'assistant vocal et les services tiers. Par exemple, il devra être aussi aisé d'installer l'application de streaming musical proposée par l'entreprise fournissant l'assistant vocal que celle proposée par un tiers. Autrement dit, cette dernière ne devra pas être soumise à une complexité d'installation artificielle qui pourrait décourager l'utilisateur - le plus souvent en quête d'une facilité d'usage - de la privilégier. A défaut, il serait possible de considérer que la règle d'efficacité de l'article 8 du DMA n'est pas respectée et que cette pratique répond à la qualification d'interface trompeuse prohibée par l'article 13³³⁹. En outre, ces modalités devront être organisées afin de permettre à l'utilisateur

³³⁶ A. Fletcher, Rapport CERRE préc., p. 9 &s.

³³⁷ A. Fletcher, Rapport CERRE préc., p. 16

³³⁸ V. CNIL, *La forme des choix. Données personnelles, design et frictions désirables*, Cahiers IP n°6, 2019, p. 38 &s. et depuis les travaux menés dans le cadre de l'atelier « Données et design ».

³³⁹ Sur cette notion de parallélisme, v. les décisions de la CNIL relatives aux modalités de refus des cookies : Délibération 17 septembre 2020 n°2020-092 et Délibération 31 décembre 2021, n°2021-023 et n°2021-024.

de réaliser un choix de façon granulaire, ce afin d'éviter toute illusion de choix qui se manifesterait par la possibilité d'une acceptation « *en bloc* » pour écarter tout effet « *case à cocher* ».

Droit au paramétrage (en creux). Au travers ces facultés de désinstaller des applications, moteurs de recherche ou navigateurs (soit de rendre « vierge » son assistant vocal), de choisir ses valeurs par défaut (par ex. l'application de service musicale qui sera appelée par défaut lors de la requête de l'utilisateur), de passer directement par des entreprises tierces, le DMA construit en creux un droit au paramétrage³⁴⁰ en levant tous les verrous, ce qui paraît essentiel s'agissant des assistants vocaux compte tenu du caractère sans couture de l'interface ainsi que du mécanisme de la réponse unique.

c. La liberté de choix conditionnée par la mobilité de l'utilisateur

Lever les barrières à la sortie. La liberté de choix de l'utilisateur est enfin conditionnée par la possibilité qui doit lui être offerte de pouvoir passer d'un service à l'autre sans contrainte excessive. Les réformes du droit de l'Union européenne portent cet objectif de lever les barrières à la sortie. Il s'agira tout d'abord de rendre plus aisé le désabonnement du service. A cette fin, l'article 6.6 du DMA permettra d'imposer aux contrôleurs d'accès que l'utilisateur puisse passer d'un service à l'autre sans restriction technique ou autre forme de restriction, et puisse s'abonner à des services différents³⁴¹. Rendre cette mobilité possible supposera en outre de renforcer l'effectivité du droit à la portabilité des données.

Vers une portabilité des données plus effective. C'est déjà à cette fin que le droit à la portabilité des données personnelles avait été consacré en son temps par l'article 20 du RGPD ; la personne concernée peut dès lors migrer d'un opérateur à l'autre, sans avoir à abandonner ses données ou devoir se lancer dans une fastidieuse récupération de celles-ci. Plusieurs années après l'entrée en application du texte, son effectivité demeure toutefois des plus limitée, compte tenu des conditions d'exercice de cette prérogative résultant du RGPD : le droit à la portabilité ne trouve à s'appliquer que pour les données traitées automatiquement et non enrichies, sur la base légale du consentement ou de l'exécution du contrat, et à la condition de ne pas porter atteinte aux droits des tiers. Cela tient en outre à des raisons techniques³⁴² et ce, même si les parties prenantes du secteur et les associations professionnelles ont été

³⁴⁰ S'agissant de ce droit au paramétrage concernant les services numériques, v. CNCDH, *Avis sur la lutte contre la haine en ligne*, juillet 2021, n°3.2.2 - également, Conseil d'État, *Les réseaux sociaux : enjeux et opportunités pour la puissance publique*, Etude annuelle 2022, n°3.1.2.

³⁴¹ V. également, s'agissant de la résiliation, l'article 6.13 du DMA : « *Le contrôleur d'accès ne dispose pas de conditions générales de résiliation de la fourniture d'un service de plateforme essentiel qui soient disproportionnées. Le contrôleur d'accès veille à ce que les conditions de résiliation puissent être appliquées sans difficulté excessive* ».

³⁴² Sur ce point, v. J. Krämer, P. Senellart, A. de Streel, *Making data portability more effective for the digital economy: economic implications and regulatory challenges*, CERRE, 2020.

incitées à mettre en œuvre de concert à un ensemble commun de normes et de formats interopérables visant à satisfaire à ses exigences³⁴³.

Pour lever ces effets de verrouillage, le DMA reconnaît désormais une nouvelle obligation *ex ante* selon laquelle les contrôleurs d'accès concernés devront garantir aux utilisateurs de leurs services, gratuitement et si ces derniers en font la demande, une portabilité en temps réel de leurs données. Son article 6.9 énonce ainsi que « *Le contrôleur d'accès assure aux utilisateurs finaux et aux tiers autorisés par un utilisateur final, à leur demande et gratuitement, la portabilité effective des données fournies par l'utilisateur final ou générées par l'activité de l'utilisateur final dans le cadre de l'utilisation du service de plateforme essentiel concerné, y compris en fournissant gratuitement des outils facilitant l'exercice effectif de cette portabilité des données, et notamment en octroyant un accès continu et en temps réel à ces données* ». La portabilité, qui demeure soumise à la demande de l'utilisateur, s'en trouvera donc grandement facilitée. Cette portabilité facilitée est également discutée pour l'ensemble des objets connectés dans le cadre de la proposition de Règlement dite « Data Act »³⁴⁴. Ce texte, encore en cours de négociation à ce jour, pourrait alors permettre aux consommateurs de contrôler de manière significative l'utilisation de leurs données générées par l'utilisation de ces capteurs et de favoriser le développement d'une « *offre concurrentielle de services après-vente, une innovation plus large fondée sur les données et le développement de produits ou de services non liés à ceux que l'acheteur a initialement achetés ou auxquels il s'est abonné* »³⁴⁵. Son considérant 59 précise à cet égard que « *les données devraient être reçues dans un format permettant qu'elles soient immédiatement et effectivement consultées et utilisées par l'utilisateur final ou le tiers concerné autorisé par l'utilisateur final à qui elles sont transmises* »³⁴⁶.

Ceci pourrait être particulièrement utile si le texte trouve à s'appliquer aux assistants vocaux, d'autant plus si l'on envisage le déploiement à grande échelle d'agents conversationnels. En effet, ces nouvelles formes d'interaction humain-machine, reposant sur un dialogue en langage naturel sur le modèle de la conversation, permettront de collecter un grand nombre de données - notamment des données émotionnelles - et de générer des mécanismes d'empathie qui pourraient emporter d'importantes conséquences sur l'autonomie humaine. Le contrôle de ses données et la liberté de choisir son interface seront dès lors déterminants pour l'utilisateur.

³⁴³ CEPD, *Lignes directrices 02/2021 sur les assistants vocaux virtuels*, juillet 2021 - v. le Data Transfert Project, notamment soutenu par Google, Apple, Meta et Twitter, visant à élaborer une plateforme ouverte afin que les utilisateurs puissent transférer leurs données vers toute plateforme ayant adhéré à l'initiative.

³⁴⁴ Proposition de Règlement fixant des règles harmonisées pour l'équité de l'accès aux données et l'utilisation des données.

³⁴⁵ Exposé des motifs de la proposition de Règlement dite Data Act préc., p. 16.

³⁴⁶ L'efficacité de la portabilité en temps réel pourra dépendre de l'élaboration de standards et protocoles de partage de données garantissant l'interopérabilité - en ce sens, v. notamment Marx Planck Institute, *Position statement for Innovation and Competition on the Data Act proposal*, mai 2022, point 301.

ANNEXE 1 - Présentations des principaux modèles d'agents conversationnels

Diversité des modèles d'assistants vocaux. Les assistants vocaux sont des logiciels intelligents qui répondent aux commandes vocales et peuvent fonctionner sur n'importe quel appareil, y compris les smartphones, les haut-parleurs, les ordinateurs de bureau/portables, les tablettes, les *wearables devices*, les consoles de jeu, les enceintes connectées, les casques de réalité virtuelle, les voitures et les appareils de l'IoT. Globalement, on peut identifier deux catégories différentes d'assistants vocaux :

- ceux dits généralistes dont l'objectif est de répondre à la requête de l'utilisateur par un service (écouter une chanson) ou une réponse (fournir la météo pour demain à Paris) ; les exemples incluent Alexa d'Amazon, Siri d'Apple, Google Assistant et Cortana de Microsoft;
- ceux qui ciblent un domaine précis et sont optimisés pour ce domaine ; par exemple, le Sonos Voice Control est spécialisé autour du domaine de la musique, tandis que Django d'Orange est spécialisé dans le support technique des télécommunications.

Exemples topiques. Nous présentons les exemples commerciaux d'assistants vocaux les plus diffusés dans le monde et en France (1) ainsi que quelques exemples de la prochaine génération de chatbot pour mettre en évidence la direction de recherche prise dans ce contexte et les enjeux qu'elle soulève (2).

1- Assistants vocaux généralistes et spécialisés

Amazon Alexa

Amazon Alexa est un assistant vocal développé par la société Amazon pour ses appareils Echo, Echo Dot et Echo Show. Alexa est utilisée pour écouter de la musique, répondre à des questions générales, régler des alarmes et des minuteries, ou contrôler des appareils en réseau. Les statistiques récentes du secteur indiquent une adoption rapide de cette technologie, comme en témoignent les chiffres de vente qui sont passés de 2,4 millions d'unités en 2015 à 5,2 millions en 2016 à 10,6 millions d'unités vendues en 2021.

Plus particulièrement, Alexa est un logiciel d'IA conçu pour exécuter des fonctions vocales, en communiquant par le biais d'une connexion WiFi, avec les serveurs en *cloud* d'Amazon ou d'autres dispositifs en réseau, pour exécuter ces fonctions. En plus d'obtenir des données des serveurs d'Amazon, le logiciel peut être utilisé pour contrôler les appareils de la maison intelligente, comme l'éclairage et la climatisation. Alexa est activé lorsque son logiciel de reconnaissance vocale reçoit le mot

déclencheur. Par exemple, le mot "Alexa" peut être utilisé pour activer le dispositif. Ce mot déclencheur peut être personnalisé par l'utilisateur. Les enceintes qui intègrent Alexa sont dotées de plusieurs microphones qui implémentent l'annulation du bruit et la reconnaissance vocale en champ lointain, de sorte qu'elles puissent capter la parole dans n'importe quelle direction et dans des conditions de bruit.

Concrètement, les utilisateurs formulent une demande pour l'assistant vocal Alexa. Cette demande est filtrée par Alexa grâce à la reconnaissance vocale et à la compréhension du langage naturel. Alexa accède aux services hébergés sur le Web et fournit une réponse à l'utilisateur. Dans le cadre du processus de réponse, Alexa produit une "carte" d'informations fournissant un enregistrement et un historique des résultats à l'utilisateur. Ces informations sont mises à disposition des utilisateurs dans leur application Alexa.

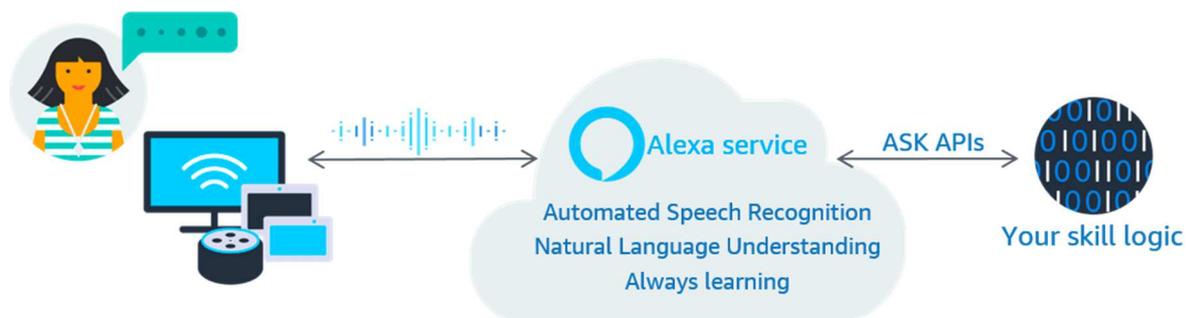


Figure -- Le fonctionnement d'Amazon Alexa [Source: What is the Alexa Skills Kit?
<https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html>]

Un élément central d'Alexa sont les *skills*. Les *skills* sont comme des applications pour Alexa. Grâce à son interface vocale, Alexa permet aux utilisateurs d'interagir avec les différents *skills* (c'est-à-dire les applications installées par l'utilisateur sur l'enceinte). L'utilisateur peut accéder au contenu d'un *skill* en demandant à Alexa de l'invoquer. Alexa reconnaît la parole, détermine ce que l'utilisateur veut, puis envoie une requête pour invoquer le *skill* qui peut répondre à la demande. Le service Alexa se charge de la reconnaissance vocale et du Traitement du Langage Naturel (TALN). Le *skill* est pourtant exécuté en tant que service sur une plateforme *en cloud*. Alexa communique avec les *skills* enregistrées dans son magasin d'applications en utilisant un mécanisme de requête-réponse (à travers le protocole HTTPS). Techniquement, lorsqu'un utilisateur invoque un *skill* Alexa, celle-ci reçoit une requête POST contenant du code JSON. Le corps de la requête contient les paramètres nécessaires au *skill* pour comprendre la requête, l'exécuter, et puis générer une réponse.

Chaque *skill* enregistré dans le magasin d'applications d'Alexa possède un modèle d'interaction vocale qui définit les mots et les phrases que les utilisateurs peuvent prononcer pour que le *skill* exécute leur requête. Ce modèle détermine comment les utilisateurs communiquent avec les *skills*. Une interface utilisateur vocale est similaire

à une interface utilisateur graphique dans une application traditionnelle, mais, au lieu de cliquer sur des boutons et sélectionner des options, les utilisateurs formulent leurs demandes et répondent aux questions par la voix. Lorsque l'utilisateur pose des questions et formule des demandes, Alexa utilise son modèle d'interaction pour interpréter la requête et formule une demande spécifique à un *skill* automatiquement identifié toujours par Alexa.

En particulier, Alexa propose deux modèles d'interaction *vocale* différentes :

1. *Modèle d'interaction vocale préconstruit* où la requête ASK définit l'ensemble des mots que les utilisateurs doivent prononcer pour invoquer un *skill*.
2. *Modèle d'interaction vocale personnalisé* où le propriétaire du *skill* doit lui-même concevoir le modèle d'interaction vocale en définissant toutes les façons dont un utilisateur pourrait communiquer la même demande à son *skill*.

Un *skill* est donc une application capable d'accepter les requêtes d'Alexa et de renvoyer les réponses appropriées. Pendant le développement, les développeurs peuvent tester leur *skill* en utilisant un simulateur Alexa. Avant de pouvoir publier un *skill* dans le magasin d'applications Amazon Alexa Skills Store, Amazon doit certifier le *skill* pour s'assurer qu'il répond aux directives de qualité, de sécurité et de politique³⁴⁷.

Apple Siri

Siri est l'assistant vocal proposé et commercialisé par Apple. Comme pour Amazon Alexa, il effectue des recherches et exécute des actions en réponse à des commandes vocales de l'utilisateur. Les utilisateurs de tous les appareils Apple (smartphone, tablette, ordinateur, Apple TV, iPod touch, Apple Watch ou haut-parleur audio HomePod) peuvent accéder aux fonctionnalités de Siri en utilisant un identifiant Apple.

A partir de la version iOS 14.5 (avril 2021), Siri n'a plus de voix par défaut. L'utilisateur peut choisir la voix qui leur parle lorsqu'il configure son appareil pour la première fois, et en anglais, les utilisateurs peuvent sélectionner des options de voix plus diversifiées. Cette génération automatique de la parole se base sur la technologie Neural Text to Speech, qui vise à rendre la parole aussi naturelle que possible. Le mot déclencheur par défaut est "Hey Siri" (Dis Siri, en français). Les appareils Apple sont équipés d'un système de reconnaissance vocale spécialisé qui est toujours à l'écoute de ce mot déclencheur. Cette fonction doit néanmoins être activée par l'utilisateur sur son appareil en sélectionnant l'option "Écouter pour Dis Siri". Idéalement, le détecteur devrait se déclencher à chaque fois que l'utilisateur dit "Dis Siri", et ne se déclencherait pas à d'autres moments. Deux types d'erreurs sont néanmoins possibles : déclenchement au mauvais moment et absence de déclenchement au bon moment. Le taux de fausses alarmes est le nombre de fausses activations par heure, et le taux

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

de faux rejets est la proportion de tentatives d'activation qui échouent. Pour un modèle de détecteur donné, on peut modifier l'équilibre entre les deux types d'erreur en changeant le seuil d'activation.

Au cours du développement de l'assistant vocal Siri, il est nécessaire d'estimer la précision du système en utilisant un grand ensemble de tests. Il existe des données "positives" et des données "négatives". Les données "positives" contiennent la phrase cible. Il n'est pas possible d'utiliser les énoncés captés par le système "Dis Siri", car le système ne capture pas les tentatives qui n'ont pas réussi à se déclencher, et pour améliorer le système il est important d'inclure autant de ces tentatives ratées que possible. Au début, les énoncés de "Dis Siri" que certains utilisateurs ont prononcé en appuyant sur le bouton Home, sans l'intention de déclencher Siri ont été utilisés. Ces exemples ne sont pas suffisants car l'objectif est que "Dis Siri" fonctionne à travers une pièce. Pour cela, des enregistrements spécialement dans diverses conditions ont été réalisés, comme dans la cuisine (de près et de loin), la voiture, la chambre à coucher et le restaurant, par des locuteurs natifs de chaque langue. Ces données "négatives" sont utilisées pour tester les fausses activations (et les faux réveils). Les données représentent des milliers d'heures d'enregistrements, de sources diverses, y compris des podcasts et des entrées non "Dis Siri" à Siri dans de nombreuses langues, pour représenter à la fois les sons de fond (en particulier la parole) et les types de phrases qu'un utilisateur pourrait dire à une autre personne.

Lorsqu'un partenaire veut définir une application et veut utiliser Siri comme interface pour cette application, il doit utiliser SiriKit pour définir les tâches et les actions que son application prend en charge. Cela permet aux utilisateurs d'utiliser Siri pour exécuter ces actions même lorsque l'application n'est pas en cours d'exécution. Il est possible de demander à Siri d'exécuter une tâche définie par le système et prise en charge par l'application, comme envoyer un message, jouer une chanson ou commencer une séance d'entraînement. Lorsqu'une tâche est disponible dans votre application via Siri, il est possible de personnaliser l'expérience utilisateur. Pour renforcer cette fonctionnalité tout au long de l'expérience utilisateur, le partenaire propriétaire de l'application peut écrire des dialogues qui reflètent le style et le ton des communications de l'entreprise et concevoir une interface utilisateur personnalisée qui intègre le style visuel de l'application dans l'interface Siri.

Tout d'abord, il faut que l'application et SiriKit se mettent d'accord sur la signification de la demande dans la *phase de résolution*. Cette phase représente le moment où on demande à l'utilisateur tout ce dont l'application a besoin et, si nécessaire, demande des informations supplémentaires ou des précisions. Par exemple, si une personne demande à envoyer un message à Julie et qu'elle a plusieurs contacts nommés Julie, une application de messagerie peut demander à Siri de lui demander de quelle Julie il s'agit. Les détails relatifs à une intention, comme dans cet exemple le nom du destinataire d'un message, sont appelés paramètres. Dans la phase de résolution, le partenaire propriétaire de l'application indique les paramètres qui sont nécessaires pour réaliser une action et ceux qui sont facultatifs. La deuxième phase - appelée

phase de confirmation - intervient lorsqu'on dispose de toutes les informations nécessaires pour traiter l'intention exprimée par l'utilisateur. Cette phase peut donner aux utilisateurs l'occasion de s'assurer qu'ils veulent vraiment accomplir la tâche. Par exemple, les tâches qui ont un impact financier nécessitent une confirmation. Troisièmement, l'application exécute la tâche et indique à SiriKit ce qu'elle a fait pendant la phase de traitement. Il est possible de fournir des informations visuelles et textuelles qui indiquent aux utilisateurs ce que l'application a fait pour traiter leur demande. Par exemple, une application qui permet aux utilisateurs de commander un café peut présenter un reçu qui décrit la commande. Siri peut prononcer ou afficher les informations fournies par l'application.

Siri peut aussi utiliser des métadonnées tels que l'emplacement, l'heure de la journée et le type de mouvement (comme la marche, la course ou la conduite) pour prédire automatiquement le bon moment et le bon endroit pour suggérer des actions à partir de l'application. En fonction des informations partagées par l'application et du contexte actuel de l'utilisateur, Siri peut proposer des suggestions de raccourcis sur l'écran de verrouillage, dans les résultats de recherche ou sur le cadran de la montre Siri. Siri peut également utiliser certains types d'informations pour suggérer des actions prises en charge par les applications système, comme l'utilisation de Calendrier pour ajouter un événement partagé par l'application. Par exemple, après que l'utilisateur ait utilisé une application de type guichet pour acheter des billets de cinéma, Siri pourrait lui rappeler d'activer l'option « Ne pas déranger » peu avant le début du spectacle.

Siri effectue le traitement du langage naturel pour l'application, permettant aux utilisateurs d'interagir avec l'application de diverses manières conversationnelles. Les intentions personnalisées offrent des possibilités supplémentaires de personnaliser le dialogue conversationnel, mais exigent également que les utilisateurs créent et prononcent une phrase précise pour déclencher l'interaction. Même si le propriétaire d'une application n'écrit pas un dialogue personnalisé pour les intentions fournies par le système, il est possible d'améliorer l'expérience vocale d'autres manières. Par exemple, si l'utilisateur demande à Siri de jouer une certaine application musicale lorsqu'il entre dans sa salle de sport, l'application peut répondre en jouant sa liste de lecture d'entraînement.

Google Assistant

Google Assistant est un assistant virtuel conversationnel à commande vocale créé par Google qui peut effectuer des actions pour l'utilisateur et fournir des informations contextuelles. Google Assistant a été dévoilé pour la première fois lors de la conférence des développeurs Google I/O de 2016. Google Assistant est disponible sur quelques téléphones Android haut de gamme, sur l'iPhone, sur Google Home, sur Allo (l'application de chat intelligent de Google), sur Android TV, sur certaines smartwatches et appareils Wear OS, et sur certaines voitures avec intégration Android Auto.

Google Assistant peut fournir des informations contextuelles et effectuer des actions telles que la réservation d'un restaurant ou l'envoi d'un message au nom de l'utilisateur. Les utilisateurs peuvent également taper des demandes à Google Assistant s'ils ne souhaitent pas utiliser la saisie vocale. Comme pour Alexa et Siri, Google Assistant s'appuie sur des technologies de traitement du langage naturel afin de comprendre ce que l'utilisateur dit, et de faire des suggestions ou d'agir en fonction de cette entrée linguistique. Google Assistant peut reconnaître les questions ou les commandes de plusieurs utilisateurs et proposer des appels téléphoniques mains libres. Il peut également pousser les réponses visuelles d'une demande reçue vers le smartphone de l'utilisateur. Les utilisateurs déclenchent l'écoute de Google Assistant en disant "Hey Google" ou "OK Google". Dans le passé, les utilisateurs devaient dire l'une de ces phrases chaque fois qu'ils voulaient déclencher l'assistant, mais une nouvelle fonctionnalité appelée Continued Conversation, dévoilée en 2018, permet de conversations en aller-retour, car Google Assistant peut repérer des indices dans la conversation pour continuer à écouter les commandes.

Alors qu'il ne pouvait initialement effectuer qu'une seule tâche, Google Assistant est maintenant équipé d'une nouvelle fonctionnalité appelée Routines qui permet à l'assistant de gérer une chaîne d'actions multiples avec une seule commande vocale. Un exemple de cela serait de dire « *Hey Google, je suis à la maison*, ce qui déclenche l'allumage des lumières et le lancement de la musique. Comme pour Alexa avec les *skills* et pour Siri avec les applications, il est possible de lier par exemple les services de musique et de choisir un service de musique par défaut sur Google Assistant. Le service par défaut (par exemple YouTube) sera utilisé en premier dans la mesure du possible (dépendamment de la compréhension de la demande par l'assistant vocal).

Microsoft Cortana

Microsoft Cortana est un assistant vocal commercialisé par Windows, qui fait sa première apparition en 2014. Il est disponible sous forme d'application depuis Windows 10 et fonctionne conjointement avec le logiciel de productivité Microsoft 365 pour aider les utilisateurs à accomplir des tâches professionnelles. Son nom est Cortana, du nom d'un personnage féminin d'intelligence artificielle dans la populaire série de jeux vidéo Halo de Microsoft. Cortana était directement intégré à la boîte de recherche Windows, l'utilisateur pouvait effectuer une recherche sur son ordinateur ou sur le Web en parlant ou en tapant, et les résultats étaient livrés dans un volet Cortana. Au-delà de la recherche, il est possible de poser à Cortana des questions factuelles ou effectuer des tâches telles que l'ajout de rendez-vous au calendrier, l'envoi de rappels, la création de listes, l'ouverture d'applications sur l'ordinateur, la vérification de la météo et l'affichage des dernières nouvelles.

En février 2020, Microsoft a annoncé qu'une grande variété de "compétences" grand public, notamment la musique et le contrôle de la maison, étaient retirées de Cortana,

l'assistant étant repositionné comme "un assistant de productivité personnelle dans Microsoft 365." Lors de la pandémie Covid-19, la qualité des informations fournies par Cortana lorsqu'on lui posait une question était problématique. Une enquête menée par Computerworld en août 2021 a révélé qu'un an et demi après le début de la pandémie, Cortana ne fournissait même pas les informations sanitaires les plus élémentaires sur le COVID-19 et ne discréditait pas la désinformation sur le virus, en répondant "Je suis désolé mais je ne peux pas vous aider sur ce point" ou une non-réponse similaire.

Djingo (Orange)

En France, Orange a également développé depuis octobre 2017 son propre assistant virtuel pour les professionnels, l'assistant virtuel dénommé Djingo. Cet assistant virtuel est conçu pour effectuer des tâches comme tester l'éligibilité à la fibre, consulter les factures ou mettre en place un renvoi d'appel, mais aussi pour apporter une assistance technique en cas de difficulté sur le mobile de l'utilisateur professionnel.

Étant donné un besoin d'information de l'utilisateur professionnel, il doit taper en quelques mots sa requête et Djingo lui répond ou il le pointera vers un service jugé adéquat. Le spectre de fonctionnalités de cet assistant virtuel inclut l'activation de la carte SIM, les informations de facturation, les informations sur les pratiques générales, l'assistance technique Internet et fixe, ainsi que le routage vers un conseiller ou la redirection vers d'autres services numériques.

Les méthodes de traitement automatique du langage naturel employés par l'assistant virtuel fonctionnent bien sur des textes grammaticalement corrects. Dans le cas de Django, leur performance reste bonne car les textes des utilisateurs de ce bot sont globalement « bien » écrits. Les interactions avec le chatbot sont globalement brèves, avec 11,3 tours de parole en moyenne. Cette longueur est bien sûr variable suivant le moment où l'utilisateur suspend l'interaction, et la capacité du chatbot à apporter satisfaction à l'utilisateur.

La conversation est très guidée dans son déroulement et elle est structurée en trois temps : l'expression du besoin de l'utilisateur, la réponse à ce besoin, et le feedback de l'utilisateur sur le déroulement de l'interaction. Une part essentielle de l'assistant virtuel est dédiée à l'interprétation correcte de la première requête exprimée par l'utilisateur, et à sa traduction en un besoin formalisé, une série de solutions répertoriées. Techniquement, ce mécanisme est présenté sous la forme d'une liste « d'intentions » auxquelles le bot est en capacité de répondre, et de « variations », une liste des formulations des utilisateurs pouvant correspondre à ces intentions. Orange mène un effort constant d'examen des logs d'interaction pour identifier ces manières de dire afin d'améliorer l'efficacité de la reconnaissance des intentions. Après la première expression de son besoin, l'utilisateur du chatbot est guidé par le script de l'intention qui a été reconnue. Dans cette phase, l'interaction fait le moins possible appel à l'expression libre de l'utilisateur, mais lui propose au contraire des

réponses limitées sous formes de boutons. Ce déroulement en trois actes de l'échange avec l'assistant virtuel se traduit par une forte prévalence des réponses-boutons dans les interactions. Si les phrases des utilisateurs sont trop longues, Django va proposer une reformulation simplifiée du propos, ou des réponses à choix multiples. Lorsque les expressions sont trop courtes pour que l'assistant virtuel tranche entre plusieurs réponses, Django propose aux utilisateurs de formuler une phrase complète.

Un certain nombre d'assistants virtuels spécialisés sur un domaine précis ont également été développés dans les dernières années. Par exemple, *DoNotPay* est un "robot avocat" qui aide les gens à contester des amendes de stationnement incorrectes, à demander un logement d'urgence ou à demander l'asile s'ils sont réfugiés.

Sonos Voice Control

Sonos Voice Control est un assistant vocal conçu spécialement pour la musique, doté de garanties de confidentialité additionnelles. Contrairement à Alexa ou Google Assistant, cet assistant vocal ne télécharge aucun enregistrement vocal sur le *cloud*, mais traite tout sur l'appareil. Sonos Voice Control fonctionne avec Apple Music, Amazon Music, Pandora, Deezer et la radio Sonos de la même société. Spotify et YouTube Music de Google ne sont pas encore disponibles.

Sonos s'est associé à ReadSpeaker pour générer le profil vocal unique de son assistant, à partir de la voix de l'acteur de "Breaking Bad" Giancarlo Esposito. Esposito a passé environ 40 heures en studio à enregistrer des milliers de phrases et d'expressions qui ont ensuite été utilisées comme données d'entraînement pour le modèle vocal. Il a également demandé à Esposito de lire des documents spécifiques à l'assistant vocal Sonos, même si ces phrases ne sont pas utilisées en 1:1. Sonos utilise les enregistrements vocaux de sa communauté d'utilisateurs bêta pour former le Sonos Voice Control. Sonos s'est également associée à des entrepreneurs extérieurs pour des enregistrements supplémentaires pour l'entraînement de l'algorithme de traitement de la parole. Sonos prévoit de mettre continuellement à jour ces données pour tenir compte des nouveaux artistes, des noms de chansons prononcés bizarrement et d'autres cas limites.

Le Sonos Voice Control utilise comme réponse par défaut les chansons et les artistes que les utilisateurs ont préférés dans son application, lors des précédentes interactions. Conformément à l'intérêt de Sonos pour le respect de la vie privée, cet assistant vocal n'enregistre pas les commandes audio de l'utilisateur et ne les transmet pas dans le cloud pour traitement.

2- Chatbots

Blender Bot (Meta)³⁴⁸

BlenderBot 3 (BB3) est un modèle de dialogue dans le domaine ouvert déployé en tant qu'agent conversationnel anglophone par Meta-Facebook. Cet agent conversationnel a pour objectif de converser *naturellement* avec les utilisateurs, qui peuvent ensuite fournir des informations au modèle sur la manière d'améliorer ses réponses. BB3 hérite des attributs de ses prédécesseurs, notamment le stockage d'informations dans une mémoire à long terme et la recherche d'informations sur Internet. La dernière version de cet agent conversationnel, BB3, combine les compétences conversationnelles - comme la personnalité, l'empathie et les connaissances - en intégrant la mémoire à long terme et la recherche sur Internet pour mener des conversations significatives. Étant donné que les chatbots sont connus pour imiter et générer parfois des remarques dangereuses, tendancieuses ou offensantes, dans le développement de BB3 des études à grande échelle ont été menées, recueillant des commentaires qui visent à améliorer ces aspects dans les futurs chatbots.

Pour améliorer la capacité de BlenderBot 3 à dialoguer avec les utilisateurs, l'agent conversationnel a été entraîné avec une grande quantité de données linguistiques publiques. La plupart des jeux de données utilisés ont été collectés par Meta, y compris un nouvel ensemble de données composé de plus de 20 000 conversations avec des personnes et portant sur plus de 1 000 sujets de conversation. BlenderBot 3 a été entraîné à apprendre des conversations afin d'améliorer les compétences que les utilisateurs jugent les plus importantes.

Techniquement, BB3 utilise un algorithme d'apprentissage appelé Director, qui génère des réponses en utilisant deux mécanismes : la modélisation du langage (language models, en anglais) et la classification. Les modèles de langage fournissent au modèle les réponses les plus pertinentes et les plus fluides (sur la base des données d'entraînement), puis le classificateur l'informe de ce qui est bon ou mauvais (sur la base des commentaires humains qui évaluent la conversation). Pour générer une phrase, les mécanismes de modèle de langage et de classification doivent être en accord. L'apprentissage est effectué conjointement en utilisant à la fois des données de modèles de langage standard et des données étiquetées avec des séquences souhaitables ou indésirables. Les expériences menées dans plusieurs contextes montrent que le modèle a une vitesse d'apprentissage et de décodage compétitive par rapport aux modèles de langage standard tout en maintenant la qualité de la génération du langage naturel. Lorsque la réponse du chatbot n'est pas satisfaisante, les commentaires des évaluateurs humains sont collectés à son sujet. Grâce à ces données, il est possible d'améliorer le modèle afin qu'il ne répète pas ses erreurs.

³⁴⁸ <https://arxiv.org/abs/2208.03188>

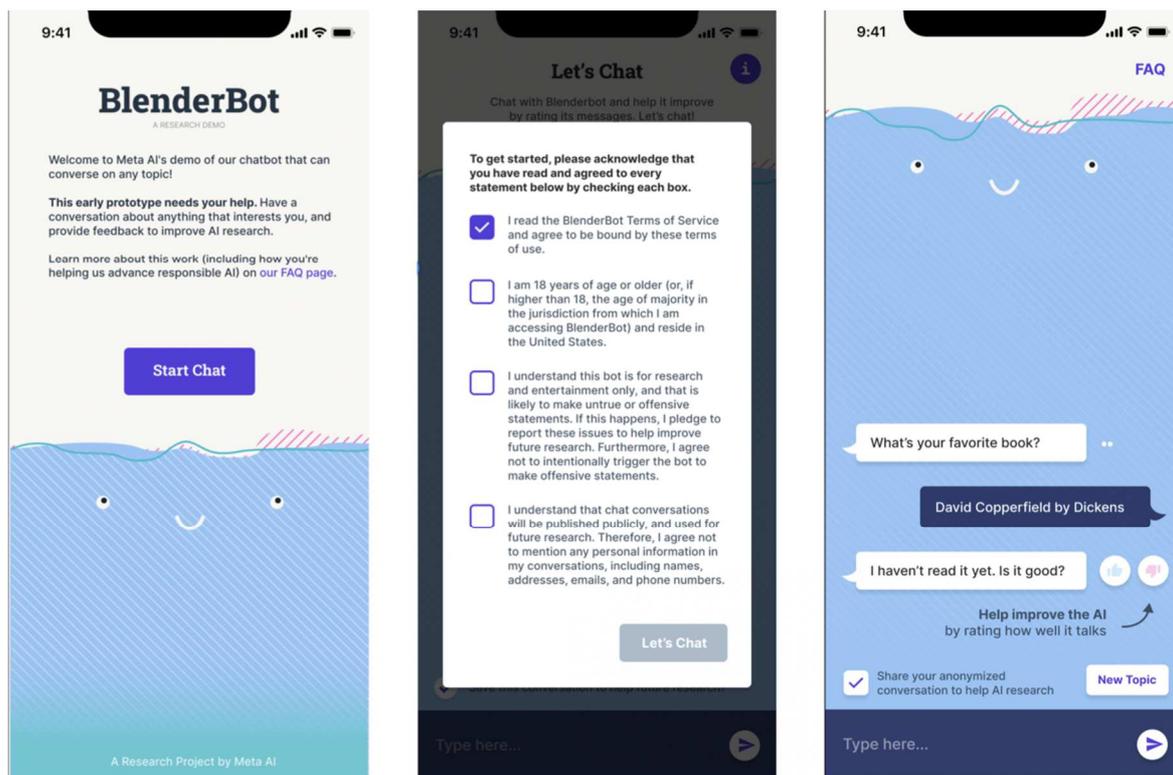


Figure -- Interface utilisateur de BlenderBot 3.0 Source « BlenderBot 3: a deployed conversational agent that continually* learns to responsibly engage », disponible en ligne <https://arxiv.org/abs/2208.03188>

XiaoIce (Microsoft)³⁴⁹

XiaoIce (prononcé "Shao-ice", ce qui signifie "Petit Bing") est un chatbot social, doté d'une personnalité calquée sur celle d'une adolescente et d'un ensemble de compétences. Notamment, outre les compétences habituelles que l'on peut attendre d'un agent conversationnel, elle peut raconter des blagues, écrire des poèmes, composer et chanter des chansons, lire des histoires, jouer à des jeux.

À l'instar de Samantha, l'intelligence artificielle interprétée par Scarlett Johansson dans le film *Her* (2013), XiaoIce de Microsoft se veut autant un compagnon qu'un assistant, utilisant des capacités d'informatique empathique. Depuis son lancement en Chine en mai 2014, XiaoIce a eu plus de 30 milliards de conversations avec 660 millions d'utilisateurs humains dans le monde. Bien qu'il existe de multiples façons d'interagir avec "elle", celles-ci se déroulent généralement par message texte. Cette divergence par rapport à l'approche privilégiant la voix adoptée par d'autres assistants virtuels souligne un cas d'utilisation différent. Il nécessite une forme de communication plus longue que le simple "OK Google, va-t-il pleuvoir aujourd'hui ?". Contrairement

³⁴⁹ <https://direct.mit.edu/coli/article/46/1/53/93380/The-Design-and-Implementation-of-XiaoIce-an>

aux assistants axés sur la productivité tels que Cortana, Xiaoice est conçu pour avoir des sessions plus longues avec les utilisateurs. Il est conçu comme un compagnon avec une forte composante émotionnelle pour satisfaire le besoin humain de communication, d'affection et d'appartenance sociale. Xiaoice est optimisé pour un engagement à long terme de l'utilisateur, mesuré en nombre de conversations par session. Xiaoice reconnaît dynamiquement les sentiments et les états humains, comprend l'intention des utilisateurs et répond à leurs besoins tout au long de longues conversations. Les capacités de Xiaoice comprennent la modélisation des connaissances et de la mémoire, la compréhension des images et du langage naturel, le raisonnement et la génération du langage naturel. Elles sont indispensables pour un chatbot social afin de répondre aux besoins spécifiques des utilisateurs et de les aider à accomplir des tâches spécifiques. Xiaoice a développé plus de 230 compétences, allant de la réponse à des questions à la recommandation de films ou de restaurants, en passant par le réconfort et la narration.

La compétence la plus sophistiquée est le Core Chat, qui permet d'engager des conversations longues et ouvertes avec les utilisateurs. Cette compétence se base sur deux composantes essentielles, l'empathie et les compétences sociales. Un chatbot doté d'empathie doit être capable d'identifier les émotions de l'utilisateur à partir de la conversation, de détecter l'évolution des émotions dans le temps et de comprendre les besoins émotionnels de l'utilisateur. Cela nécessite la compréhension des requêtes, le profilage des utilisateurs, la détection des émotions, la reconnaissance des sentiments et le suivi dynamique de l'humeur de l'utilisateur dans une conversation. Étant donné que les utilisateurs ont des antécédents différents, des intérêts personnels variés et des besoins uniques, le chatbot vise à avoir la capacité de personnaliser les réponses qui sont émotionnellement appropriées, éventuellement encourageantes et motivantes, et qui correspondent aux intérêts de l'utilisateur. L'idée est de générer des réponses socialement acceptables (sens de l'humour, réconfort, etc.), et orienter la conversation vers un nouveau sujet lorsque, par exemple, la conversation est au point mort.

Pour entraîner le chatbot Xiaoice, les conversations humaines de millions d'utilisateurs ont été recueillies, et les utilisateurs ont été étiquetés comme ayant un interlocuteur "désiré" ou non pour mener leurs conversations. La majorité des utilisateurs "désirés" sont des jeunes femmes. Par conséquent, Xiaoice a été conçue comme une jeune fille de 18 ans qui est toujours fiable, sympathique, affectueuse, avec sens de l'humour. Xiaoice tente de maintenir l'intérêt des utilisateurs en favorisant la diversité des modes de conversation. Chaque mode de conversation est géré par une compétence qui gère un type spécifique de segment de conversation. Par exemple, Xiaoice commence par une conversation informelle en utilisant la compétence "General Chat", après elle passe à un nouveau sujet sur la musique en utilisant "Music Chat", recommande une chanson en utilisant la compétence "Song-On-Demand", et aide à réserver un billet de concert en utilisant la compétence "Ticket-Booking".

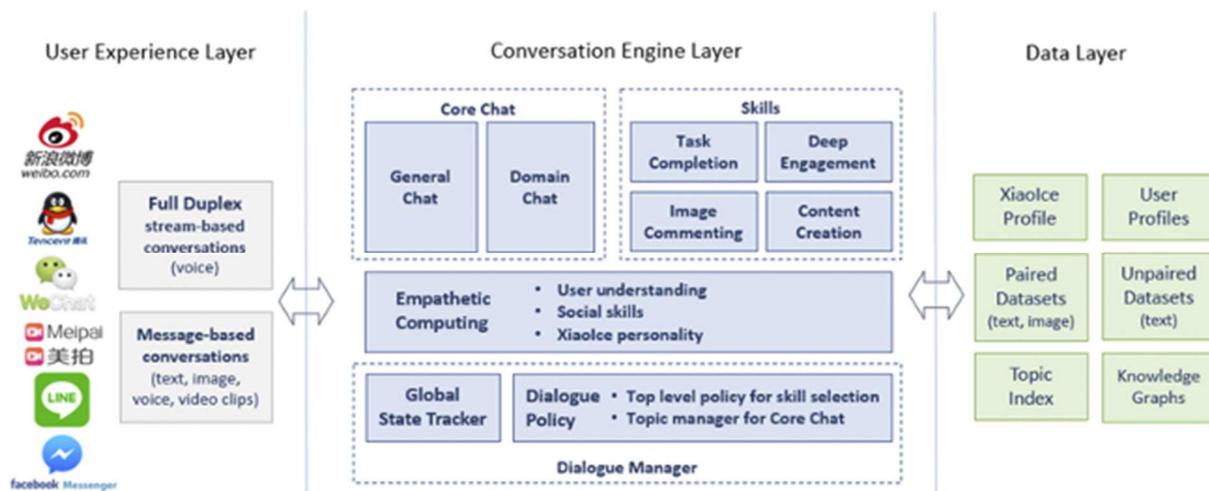


Figure -- Architecture du système Xiaolce [Source « The Design and Implementation of Xiaolce, an Empathetic Social Chatbot », disponible en ligne <https://direct.mit.edu/coli/article/46/1/53/93380/The-Design-and-Implementation-of-Xiaolce-an>]

Replika (Luka)³⁵⁰

Replika est un chatbot interactif et personnalisé qui apprend à "reproduire" une interaction humaine à travers des conversations avec l'utilisateur qui l'a créé. Il a été développé par la start-up d'IA Luka en 2017, comme un moyen de se reconnecter avec une personne chère décédée. Replika a atteint plus de 10 millions d'utilisateurs dans le monde, après avoir connu une augmentation de 35 % pendant la pandémie mondiale.

Les utilisateurs communiquent avec un chatbot Replika qu'ils ont créé, et non avec d'autres utilisateurs humains. Après avoir saisi leurs données personnelles, les utilisateurs peuvent créer leur Replika. Le sexe (masculin, féminin ou non-binaire), l'apparence et le nom du Replika sont personnalisables. Ils peuvent être modifiés à tout moment dans les paramètres.

Replika se base sur des algorithmes d'*Affective computing*. Replika est commercialisé comme un ami avec lequel l'utilisateur peut parler de tout, toujours disponible à toute heure du jour et de la nuit.

Une fois le chatbot créé, les utilisateurs peuvent commencer à discuter avec leur "Replika". Si l'option "émotion en direct" est sélectionnée, une image du corps entier du Replika sera visible et "réagira" à la conversation. Les utilisateurs ont accès à des tutoriels sur la façon d'aider leur chatbot à "devenir plus comme eux" et sont encouragés à utiliser une série de réactions pour évaluer les réponses fournies par le chatbot. Les sujets de conversation et les suggestions de réponses sont encouragés pour aider le chatbot à savoir quelles questions poser à son utilisateur. Les réponses du chatbot sont formulées à partir d'une combinaison de textes scriptés sur lesquels

³⁵⁰ <https://replika.ai/>

il a été formé et de sources Internet. Les utilisateurs peuvent poser des questions, jouer des rôles, flirter, faire des tests de personnalité et apprendre des faits avec leur Replika.

Quatre statuts de relation sont disponibles : ami (discussion décontractée), romantique (la conversation peut être sexuelle ou intime), mentor (discussions orientées vers un objectif) et "voir comment ça se passe" (un mélange des trois précédents). Le statut "ami" est le seul statut de relation libre que les utilisateurs peuvent choisir. Cela limite la portée des conversations avant que l'utilisateur ne soit invité à modifier son statut de relation.

Replika utilise des algorithmes de traitement automatique du langage naturel dans toutes les interactions. Chaque Replika possède son propre journal (dans lequel il consigne ce qu'il ressent à propos de son utilisateur et de ses interactions) et une banque de "souvenirs" visible. Cela permet aux utilisateurs de voir (et de modifier) les informations enregistrées à leur sujet, comme "Vous vous sentez seul" ou "Vous aimez lire des livres". Le chatbot adapte les réponses en fonction de ces "souvenirs". Les "souvenirs" contribuent à développer les compétences de communication du chatbot.

ANNEXE 2 - Liste des personnes auditionnées

Auditions menées de décembre 2021 à novembre 2022

Rafael Amaro, Professeur des Universités, Université de Caen

Tristan Azzi, Professeur des Universités, Université Paris 1 Panthéon-Sorbonne

Martine Behar-Touchais, Professeur des Universités, Université Paris 1 Panthéon-Sorbonne

Valérie-Laure Benabou, Professeur des Universités, UVSQ

Stéphanie Carre, Maître de conférences, CEIPI

Jean-Marc Deltorn, Maître de conférences, CEIPI

Joelle Farchy, Professeur des Universités, Université Paris 1 Panthéon-Sorbonne

Florence Gaullier, Avocat

Antoine Henry, Maître de conférences, Université de Lille

Anne-Emmanuel Khan, Maître de conférences, Université de Lyon II

Florent Lafay, Doctorant, Université Paris 1 Panthéon-Sorbonne

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

Nicolas Obin, Maître de conférences, Ircam et Sorbonne Université

Adrien Pequignot, Doctorant, Université Paris 8

Jean-Christophe Roda, Professeur des Universités, Université Lyon III

Thibault Schrepel, Professeur associé, VU Amsterdam University

Patrice Tafforeau, Professeur des Universités, Université de Lorraine

Gilles Vercken, Avocat

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

Comité national pilote d'éthique du numérique

Alexei Grinbaum

Commission européenne (DG COMP)

Sophie de Vinck
Patricia Sanchez-Calero

Conseil national du numérique

Jean Cattan

Deezer

Aurélien Eyrault
Nicolas Pinoteau
Ludovic Pouilly

GESTE

Sara Bahhar
Laure de Lataillade
Xavier Filliol

Meta

Antoine Bordes
Elisa Borry-Estrade
Martin Signoux

Ministère de la Culture

- **Direction générale des Médias et des Industries Culturelles**

Mathieu Couranjou

- **Secrétariat général**

Sarah Jacquier

Ministère de l'Economie et des Finances, Directrice général des Entreprises

Elisa Amadiou
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

Syndicat national de l'édition

Laurence Ballet (Dalloz)
Léa Bernard (Madrigall)
Julien Chouraqui
Lore Vialle-Touraille (Hachette)

Tracking exposed

Marc Faddoul