

Louvain School of Management

L'état d'adoption de l'assistant vocal en fonction de sa valeur ajoutée apportée à trois cas d'usage majeurs :

L'aide à la conduite, l'apprentissage d'une langue étrangère, les services de secours à distance.

Auteures : Pauline Fagneray & Alice Fiedelers

Promoteur : Bernard Paque

Année académique 2020-2021

Mémoire recherche en vue d'obtenir le titre de Master en Sciences de Gestion, à finalité spécialisée

Horaire de jour

Réfléchir, rédiger, remettre en question ne se fait pas seul.

Nous tenons à remercier toutes les personnes ayant contribué directement ou indirectement à l'élaboration de ce mémoire. Nous souhaitons exprimer notre reconnaissance pour toutes les discussions, commentaires et conseils enrichissants que nous avons reçus.

Nos remerciements les plus sincères vont vers Monsieur Bernard Paque, notre promoteur qui nous a aidé, prodigué moult conseils et a toujours répondu présent lors de nos sollicitations.

Nous souhaitons également remercier tous les experts que nous avons interrogés. Ceux-ci ont accepté de nous consacrer un peu de leur temps pour répondre à nos questions et ont chacun montré leur intérêt envers notre sujet. Leurs informations précieuses nous ont permis de mieux comprendre la valeur des assistants vocaux.

Sans oublier, nos relecteurs, Sabrina Fagneray, Valérie Julien et Nathalie Riboux, pour leur temps et leur aide consacrés à la relecture de notre mémoire.

TABLES DES ILLUSTRATIONS

Figure 1: Évolution des interfaces en 10 ans.....	6
Figure 2: Adoption de l'assistant vocal par pays sur la courbe d'adoption de l'innovation de Rogers ..	9
Figure 3: Nombre d'assistants vocaux numériques dans le monde de 2019 à 2024 (en milliards).....	10
Figure 4: Évolution de l'assistant virtuel sur la courbe de Gartner, de 2017 à 2019.....	10
Figure 5: Radar d'impact des technologies émergentes	12
Figure 6 : Comparaison des taux d'adoption de plusieurs technologies aux USA.....	15
Figure 7: Diagramme de Venn de l'intelligence artificielle	24
Figure 8: Fonctionnement de l'assistant vocal, de la reconnaissance vocale à la synthèse vocale	30
Figure 9: Utilisation de l'assistant vocal sur les différents terminaux	39
Figure 10: Courbe de Gartner pour les technologies émergentes en 2017	123
Figure 11: Courbe de Gartner pour les technologies émergentes en 2018.....	123
Figure 12: Courbe de Gartner pour l'intelligence artificielle en 2019	124

TABLE DES MATIERES

INTRODUCTION.....	1
CHAPITRE 1: LA QUESTION DE RECHERCHE	4
CHAPITRE 2: REVUE DE LITTÉRATURE	5
2.1 MISE EN CONTEXTE.....	5
2.1.1 <i>Évolution de l'Interaction Homme-Machine</i>	5
2.1.2 <i>État du marché des assistants vocaux</i>	7
2.1.3 <i>Besoin à l'égard des nouvelles technologies</i>	10
2.1.4 <i>Adoption des assistants vocaux</i>	11
2.2 L'INTELLIGENCE ARTIFICIELLE	16
2.2.1 <i>Définition</i>	16
2.2.2 <i>Les algorithmes</i>	18
2.2.3 <i>Historique et évolution</i>	19
2.2.3.1 Naissance de l'intelligence artificielle.....	19
2.2.3.2 Premières inventions d'intelligence artificielle	21
2.2.3.3 Hiver de l'Intelligence Artificielle (AI Winter).....	21
2.2.3.4 La renaissance de l'Intelligence Artificielle	21
2.2.3.5 Les algorithmes d'apprentissage	22
2.2.3.6 L'Intelligence Artificielle non symbolique	22
2.2.3.6.1 L'apprentissage-machine/ automatique (Machine learning).....	22
2.2.3.6.2 L'apprentissage profond (Deep learning)	24
2.2.3.7 Big data.....	25
2.2.3.8 Le futur de l'intelligence artificielle.....	26
2.3 LES ASSISTANTS VOCAUX	27
2.3.1 <i>Définition</i>	27
2.3.2 <i>Historique et évolution</i>	28
2.3.3 <i>Le fonctionnement d'un assistant vocal</i>	30
2.3.3.1 Reconnaissance vocale (Speech-to-text)	31
2.3.3.2 Le langage naturel	32
2.3.3.3 Traitement Automatique du Langage Naturel	32
2.3.3.3.1 Compréhension du langage naturel.....	33
2.3.3.3.2 La génération en langage naturel	34
2.3.3.3.3 L'apprentissage automatique appliqué au TALN	34
2.3.3.4 Synthèse vocale (Text-to-Speech)	34
2.3.3.5 Phonologie.....	35

2.3.4	<i>Limites dans le fonctionnement de l'assistant vocal</i>	35
2.3.4.1	Limitations du mode parole.....	35
2.3.4.2	Erreurs de sémantique	36
2.3.4.3	Endpointing	37
2.3.4.4	Pollution sonore.....	37
2.3.5	<i>Le marché des assistants vocaux</i>	37
2.3.5.1	Les terminaux dotés d'un assistant vocal	38
2.3.5.2	Les principaux usages.....	40
2.3.5.3	Les cadres d'utilisation.....	41
2.3.6	<i>Sécurité et confidentialité</i>	42
2.3.7	<i>Le futur des assistants vocaux</i>	43
CHAPITRE 3: MÉTHODOLOGIE		45
3.1	CHOIX DU SUJET	45
3.2	CHOIX DE L'ÉTUDE.....	45
3.3	DESCRIPTION DU DESIGN DE LA RECHERCHE	46
3.4	ANALYSE DE DONNÉES.....	47
CHAPITRE 4: PARTIE EMPIRIQUE		49
4.1	CAS D'USAGE 1 : L'AIDE À LA CONDUITE	49
4.1.1	<i>Description du cas</i>	49
4.1.2	<i>Hypothèse de recherche</i>	50
4.1.3	<i>Analyse scientifique sur base des interviews</i>	51
4.1.3.1	Facteurs clés de succès (Offre).....	51
4.1.3.2	Risques	53
4.1.3.3	Adoption de l'assistant vocal (Demande).....	54
4.1.3.3.1	Facteurs clés d'adoption	54
4.1.3.3.2	Facteurs limitant l'adoption	55
4.1.3.3.3	Adoption de technologie embarquée ou externe.....	57
4.1.3.3.4	Adoption de technologie développée par les constructeurs ou acteurs technologiques	59
4.1.3.3.5	Adoption de fonctionnalités spécifiques ou génériques.....	61
4.1.3.4	Prédictions futures.....	62
4.1.4	<i>Conclusion du cas d'usage</i>	64
4.2	CAS D'USAGE 2 : APPRENTISSAGE D'UNE LANGUE ÉTRANGÈRE.....	66
4.2.1	<i>Description du cas</i>	66
4.2.2	<i>Hypothèse de recherche</i>	67
4.2.3	<i>Analyse scientifique sur base des interviews</i>	67

4.2.3.1	Facteurs clés de succès (Offre).....	68
4.2.3.2	Adoption de l'assistant vocal (Demande).....	70
4.2.3.2.1	Facteurs clés d'adoption	70
4.2.3.2.2	Facteurs limitant l'adoption	72
4.2.3.2.3	Adoption d'une technologie complémentaire au professeur ou d'une application indépendante	74
4.2.3.2.4	Adoption d'une technologie développée par les grands acteurs ou des startups.....	76
4.2.3.3	Prédictions futures.....	77
4.2.4	<i>Conclusion du cas d'usage</i>	79
4.3	CAS D'USAGE 3 : SERVICES DE SECOURS À DISTANCE	80
4.3.1	<i>Description du cas</i>	80
4.3.2	<i>Hypothèse de recherche</i>	81
4.3.3	<i>Analyse scientifique sur base des interviews</i>	81
4.3.3.1	Facteurs clés de succès (Offre).....	82
4.3.3.2	Risques	84
4.3.3.3	Adoption de l'assistant vocal (Demande).....	85
4.3.3.3.1	Facteurs clés d'adoption	85
4.3.3.3.2	Facteurs limitant l'adoption.....	87
4.3.3.3.3	Adoption de technologie développée par les grands acteurs ou des startups.....	88
4.3.3.4	Prédictions futures.....	90
4.3.4	<i>Conclusion du cas d'usage</i>	91
	CONCLUSION	93
	LIMITES À LA RECHERCHE.....	96
	SUGGESTIONS POUR RECHERCHES FUTURES	97
	POINTS DE VUE PERSONNELS.....	98
	BIBLIOGRAPHIE	100
	ANNEXES	115
	ANNEXE 1: GLOSSAIRE	115
	ANNEXE 2 : HYPE CYCLE DE GARTNER INCLUANT LES ASSISTANTS VIRTUELS	123
	ANNEXE 3 : EXPERTS CONTACTÉS	125
	ANNEXE 4 : GUIDES D'ENTRETIEN.....	129
	<i>Lettre d'introduction</i>	129
	<i>Cas d'usage n°1 : Aide à la conduite</i>	130
	<i>Cas d'usage n°2 : Apprentissage d'une langue étrangère</i>	132
	<i>Cas d'usage n°3 : Services de secours à distance</i>	134
	ANNEXE 5 : MATRICE DE SYNTHÉTISATION DES INTERVIEWS.....	136

<i>Cas d'usage n°1 : Aide à la conduite</i>	136
<i>Cas d'usage n°2 : Apprentissage d'une langue étrangère</i>	140
<i>Cas d'usage n°3 : Services de secours à distance</i>	143
ANNEXE 6: TRANSCRIPTION DES INTERVIEWS D'EXPERTS	146
<i>Cas d'usage n°1 : Aide à la conduite</i>	148
1. Interview de Oxana Gouliaéva	148
2. Interview de Ariane Nabeth-Halber.....	158
3. Interview de Jean-François Bonastre	174
4. Interview de Soledad Bedu.....	185
<i>Cas d'usage n°2 : Apprentissage d'une langue étrangère</i>	202
1. Interview François-Xavier Fievez	202
2. Interview Robin Guérit.....	214
3. Interview de Maria Perillo.....	228
<i>Cas d'usage n°3 : Services de secours à distance</i>	238
1. Interview de Axel Legay	238
2. Interview de Ahmed Bouzid.....	251
3. Interview de Stanley Goffinet.....	264

INTRODUCTION

« D’abord une pulsation, ensuite un souffle, puis un cri et enfin une voix. D’abord des babillages, ensuite quelques mots, puis des phrases et enfin une pensée, partagée ou polarisante mais force élémentaire d’une société dans toutes ses composantes. La voix est un marqueur de maturité de notre humanité comme elle le devient pour les objets. » (Dosquet et al., 2020, p.5).

Depuis quelques années, nous entrons dans une ère où le monde est de plus en plus connecté. Les technologies de l’intelligence artificielle (ci-après IA¹) s’immiscent progressivement dans notre quotidien, une évolution qui est encouragée par le développement de l’Internet des Objets². Par ailleurs, les appareils autonomes acquièrent davantage d’intelligence dans leur manière d’interagir avec les humains. De nouvelles avancées prennent actuellement en popularité, avec, entre autres, la technologie capable de reconnaître le langage naturel humain (Polyakov et al., 2018).

En effet, l’approfondissement de connaissances dans ce domaine est susceptible de conduire à de nouveaux modes de communication naturels entre la machine et l’humain, où la machine va apprendre à maîtriser et s’adapter au langage humain pour finalement interagir avec lui (Polyakov et al., 2018). L’un de ces principaux, n’est autre que l’assistant vocal.

Le marché des assistants vocaux est au cœur de l’évolution technologique et la vie quotidienne de tous prend une toute autre dimension à l’égard de cette nouvelle tendance. Cette innovation pouvait paraître illusoire auparavant mais est actuellement une réalité omniprésente. Comme le mentionne la citation précitée, la naissance de la voix chez une personne révèle une matrice de l’humanité, tout comme elle le devient pour les objets. En effet, ce concept s’immisce dans les parties prenantes qui affectent de manière considérable la société actuelle, à savoir, entre autres, les voitures, les lieux de travaux ou encore les domiciles (CNIL, 2020).

Le marché des assistants vocaux est en plein essor et va indubitablement constituer un nombre considérable d’opportunités futures. Cette nouveauté découle d’une évolution continue de technologies. Afin de comprendre au mieux ce nouveau marché, il sera dès lors nécessaire

¹ IA, voir glossaire n°23.

² Internet des Objets, voir glossaire n°26.

d'analyser dans un premier temps les théories que recouvre l'intelligence artificielle afin d'appréhender, dans un second temps, l'avènement des assistants vocaux. Une mise en contexte sera préalablement nécessaire afin d'avoir une vision généraliste de la situation avant d'aborder le sujet dans sa profondeur.

La transition d'une interface "humain-machine" à une interface d'interaction vocale représente un changement de paradigme. Il sera primordial de connaître la façon dont une innovation telle que l'assistant vocal a vu le jour. Pour ce faire, il sera intéressant de consulter les courbes d'évolutions des tendances de Gartner afin d'observer la position des assistants vocaux et d'évaluer leur potentielle future progression. Par conséquent, la question d'adoption de cette innovation est davantage importante. En effet, les assistants vocaux vont modifier la façon d'interagir avec les machines et dès lors créer aussi bien des facteurs favorisant que limitant l'adoption de cette nouvelle technologie vocale. Ces facteurs seront intéressants à analyser afin de juger s'il s'agit d'une demande latente ou d'une demande existante des utilisateurs pour cette innovation.

La partie de la revue de littérature sur l'intelligence artificielle couvrira un historique des éléments qui ont contribué à la venue de l'intelligence artificielle, ainsi que des spécificités techniques importantes à son développement. Dans la partie consacrée aux assistants vocaux, constituant notre sujet principal, nous exposerons l'évolution de l'interface vocale de manière à obtenir une vue d'ensemble de la genèse de ceux-ci. Nous continuerons en détaillant son mode de fonctionnement technique afin de comprendre le fondement de l'interaction vocale entre une machine et un humain. Les problèmes qu'encourt ce fonctionnement sont ensuite mis en exergue puisqu'ils constituent encore de nombreux freins à l'adoption des assistants vocaux. Enfin, nous poursuivrons sur l'état du marché des assistants vocaux en évoquant les différents terminaux sur lesquels ils sont utilisés, leurs principaux usages ainsi que les secteurs qui utilisent ces technologies. Nous terminerons en abordant les problèmes de sécurité et de confidentialité que la technologie vocale peut susciter.

En effet, bien qu'omniprésentes, ces technologies ne nous sont pas toutes d'une aide fondamentale en termes de sécurité ou de valeur économique. La problématique de savoir si l'essor des assistants vocaux va constituer une valeur ajoutée dans plusieurs domaines se fait ressentir. C'est pourquoi ce mémoire a pour but d'analyser plusieurs cas pratiques majeurs dans

lesquels les assistants vocaux s'avèrent être efficaces et aptes à l'adoption par les consommateurs. Ces cas pratiques que nous avons choisis relèvent des parties prenantes susmentionnées, à savoir les voitures, les lieux de travaux ou encore les domiciles. Les cas d'usage relatifs à ces secteurs sont respectivement l'aide à la conduite, l'apprentissage d'une langue et les services de secours à distance. Nous expliquerons la raison de ce sujet et la manière dont nous l'avons approché dans le chapitre suivant.

Enfin, la partie qualitative de notre travail fera office de partie pratique. Dans celle-ci, nous développerons les trois cas d'usage précédemment cités. La description de ces cas débouchera sur trois hypothèses respectives. Nous constituerons une base de données, à travers des interviews d'experts, afin d'être en mesure d'accepter ou de réfuter les hypothèses. À dessein de conduire ces interviews, nous développerons trois guides d'entretien relatifs à chaque cas d'usage, à destination des experts. Nous conduirons des interviews semi-directives. Grâce à une rigoureuse analyse, ces entretiens permettront de répondre à nos hypothèses de recherche. S'en suivra une discussion au travers de laquelle nous répondrons à notre question de recherche. Ensuite, viendra la conclusion et finalement, les limites de notre travail seront évoquées. Nous énoncerons également quelques suggestions destinées à de potentiels futurs travaux qui viendront améliorer et/ou compléter nos recherches. Nous finirons par émettre notre point de vue quant au potentiel futur que détiennent les assistants vocaux.

CHAPITRE 1: LA QUESTION DE RECHERCHE

Comme nous l'avons vu précédemment, les assistants vocaux investissent plusieurs secteurs. L'avènement des assistants vocaux dans les voitures, les lieux de travail et les domiciles ont attiré toute notre attention. Nous nous sommes questionnées sur le fait que ce type de technologie apporterait du positif sur le long terme ou non. Nous avons alors réfléchi à différents cas d'usage que pouvaient recouvrir ces trois secteurs et sommes parvenues à une liste d'idées où l'assistant vocal pouvait, selon nos recherches, apporter une quelconque valeur ajoutée. Parmi celle-ci, nous avons envisagé les cas d'usage suivants : l'aide à la conduite, les voitures autonomes pour l'accompagnement d'enfants seuls, l'apprentissage d'une langue étrangère, la traduction, les services de secours à distance, et les soins de santé.

Nous avons réalisé que les recherches autour d'une telle problématique devaient être délimitées. Ces cas d'usage nous ont semblé intéressants afin de poser un cadre à notre travail. L'étude de tous les cas mentionnés ci-dessus aurait nécessité un travail d'une grande ampleur. En raison de certaines limites, nous avons retenu trois de ces cas pratiques. Les cas sélectionnés sont l'aide à la conduite, l'apprentissage d'une langue étrangère ainsi que les services de secours à distance.

Ces raisonnements, couplés à des discussions avec notre promoteur, monsieur Paque, nous ont amené à la décision de traiter le sujet suivant : *« L'état d'adoption de l'assistant vocal en fonction de sa valeur ajoutée apportée à trois cas d'usage majeurs : Aide à la conduite, apprentissage d'une langue étrangère, services de secours à distance. »*

Dans le but de répondre à cette problématique, nous serons amenées à développer des hypothèses intrinsèques à chaque cas d'usage. Celles-ci seront élaborées dans le courant de notre partie empirique.

CHAPITRE 2: REVUE DE LITTÉRATURE

À partir de ce chapitre, nous allons rédiger une revue de littérature actuelle, qui fait le point sur les deux concepts majeurs de ce mémoire, à savoir : les assistants vocaux et l'intelligence artificielle (IA), sans quoi ils n'auraient pas vu le jour. Cette revue porte également sur les évolutions qui ont marqué l'histoire de ces deux secteurs.

La rédaction de cette revue de littérature nous permettra de connaître davantage ces deux notions pour, ensuite, affiner notre problématique ainsi que notre question de recherche. Finalement, nous serons en mesure de créer un guide d'entretien contenant les questions à adresser aux experts lors des interviews.

Le présent mémoire n'a pas la prétention d'énoncer les mathématiques et la science exacte derrière les algorithmes³ d'intelligence artificielle ou les assistants vocaux. Cependant, nous tenterons de développer un tant soit peu leur fonctionnement.

2.1 Mise en contexte

Dans cette section, il convient de s'interroger sur la place que prennent les technologies vocales, qu'elles soient utilisées de manière individuelle ou complémentaire avec d'autres modes de communication. C'est pourquoi nous ferons le point sur l'évolution des interfaces et moyens d'interactions entre l'humain et la machine. Nous nous pencherons également sur quelques chiffres clés afin d'établir l'état des lieux du marché des assistants vocaux. Cela sera utile afin de comparer le niveau d'adoption dans les différentes régions du monde. Ensuite, nous analyserons le besoin à l'égard de cette nouvelle technologie. Finalement, nous nous intéresserons aux prédictions futures liées à l'adoption de cette technologie émergente.

2.1.1 Évolution de l'Interaction Homme-Machine

Depuis quelques années, l'assistant vocal a la cote. Bien qu'il révolutionne notre façon d'interagir avec les interfaces, il n'est pas le premier moyen mis en œuvre. « *Cliquer, toucher*

³ Algorithme, voir glossaire n°3.

et maintenant parler », tel est le chemin que nous avons emprunté dans l'évolution de la navigation (Abela, 2019).

La question de l'Interaction Homme-Machine (IHM⁴) fait débat depuis que les ordinateurs graphiques existent. Les recherches ont débuté par les travaux de Ivan Sutherland. Au début des années 1960, il a développé le SketchPad, le précurseur de la tablette tactile actuelle. Cette première interface graphique permettait de créer des diagrammes, plans etc., à l'aide d'un crayon optique (Beaudouin, 2016).

Ensuite, les premiers assistants personnels, plus connus sous l'acronyme PDA⁵ (Personal Digital Assistant) ont vu le jour en 1991. Cette innovation a marqué l'ère de l'informatique ambiante. Au fur et à mesure, l'écran s'est donc aplani et a laissé place à des interfaces tactiles sur certains appareils, les plus communs étant les écrans de smartphone (Calvary & Coutaz, 2007). En effet, l'utilisation de l'écran demandait certaines connaissances. Le tactile était plus intuitif, et la voix, quant à elle, devient un nouveau standard. L'interface vocale connaît du succès car elle assure une universalité naturelle. Comme illustré ci-dessous (voir figure 1), la technologie élimine une couche de complexité tous les dix ans pour toucher une population toujours plus étendue. La voix est donc une discipline qui supprime une difficulté supplémentaire de la technologie (Dosquet et al., 2020, p.15).

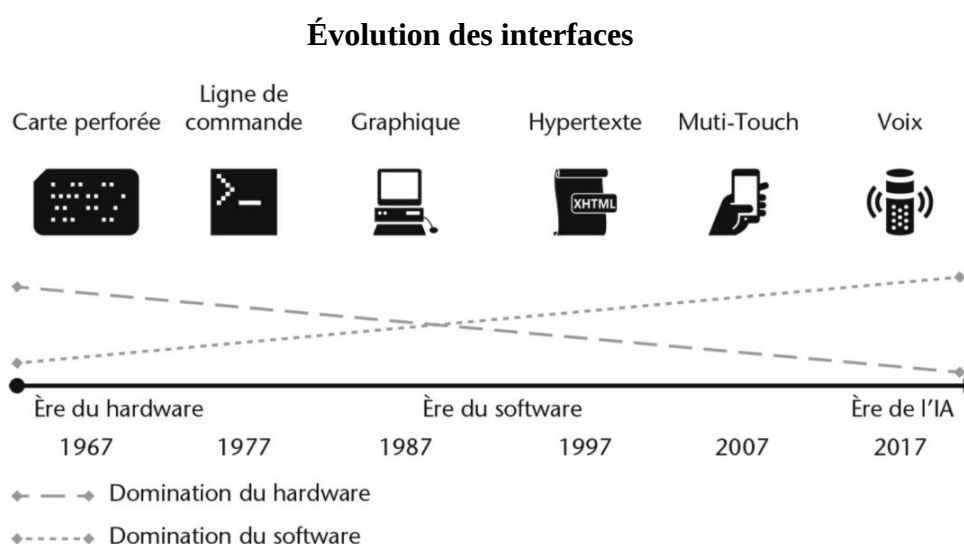


Figure 1: Évolution des interfaces en 10 ans
(Source: Dosquet et al., 2020, p.16)

⁴ Interaction Homme-Machine, voir glossaire n°25.

⁵ PDA, voir glossaire n°33.

Ces dernières années, le canal auditif est devenu très populaire. Il offre l'opportunité, que les autres canaux ne peuvent pas prétendre offrir, de faire des recherches et de se divertir sans l'usage des mains. Plusieurs facteurs semblent déterminants dans le mode « parole ». Les avantages de ce mode semblent surpasser ceux proposés par les précédents. En effet, dans le modèle humain, la parole possède cet aspect *naturel*. Enseignée dès le plus jeune âge, la parole est le mode de communication le plus naturel entre êtres humains, contrairement à l'utilisation du clavier. Ensuite, la parole étant spontanée, elle possède la caractéristique de *rapidité* et d'*efficacité*. À l'oral, nous pouvons débiter environ 200 mots par minute, contre un débit maximum de 60 mots par minutes lors de la frappe au clavier (Dosquet et al., 2020, p.13 ; Minker & Néel, 2002).

La parole étant naturelle et spontanée, l'interaction devrait également être humanisée. Il est alors essentiel d'attribuer à l'assistant vocal un ton, une personnalité et un vocabulaire adapté, favorisant l'intégration accrue de ces technologies dans notre quotidien. Par conséquent, ces technologies deviennent de plus en plus omniprésentes (Dosquet et al., 2020, p.13 & 88). La miniaturisation⁷ est un autre phénomène qui se développe en parallèle et qui favorise l'intégration accrue de ces technologies dans notre quotidien.

2.1.2 État du marché des assistants vocaux

Afin de réaliser un état des lieux du marché des assistants vocaux, nous allons nous intéresser à quelques chiffres. En plus des chiffres relatifs à la Belgique, nous analyserons également plusieurs régions du monde dont la France, en tant que pays limitrophe, afin d'établir une comparaison. Dans un souci prévisionnel, nous nous intéresserons aux données des USA, qui ont généralement une longueur d'avance en matière d'avancées technologiques. Cela pourrait nous donner quelques potentielles prédictions sur la façon dont le marché des assistants vocaux évoluera en Europe au cours des prochaines années. En effet, selon Dosquet et al. (2020, p.25), les États-Unis auraient une avance de cinq années sur la tendance que suit l'Europe.

Nous centraliserons cette analyse sur les usages d'assistants vocaux de manière générique, c'est-à-dire l'ensemble des dispositifs pouvant être dotés d'un assistant vocal. En effet, bien

⁷ La miniaturisation est « la création de produits mécaniques, optiques ou électroniques et de leurs dispositifs à des échelles de plus en plus petites » (Wikipedia, 2020).

que les enceintes connectées⁸ soient uniquement conçues à cet effet, les assistants vocaux peuvent également être utilisés sur différents terminaux (voir section 2.3.5.1).

« *Le vocal, un marché jeune mais qui «voix haut»* » (Roland Berger, 2019).

Comme l'entend la citation ci-dessus, le marché des assistants vocaux n'est pas encore mature mais semble prometteur. Le graphique réalisé par Voicebot en 2018 compare l'adoption de cette technologie au sein de différents pays, à l'aide du cycle d'adoption proposé par Everett Rogers⁹. Son modèle décrit le cycle de vie de l'adoption d'une technologie qui se compose de cinq parties, auxquelles il associe différents groupes de consommateurs correspondants. La courbe en cloche suit une certaine distribution et chaque groupe de consommateurs représente un pourcentage de la population, à savoir : les innovateurs (2,5%), les adoptants précoces (13,5%), la majorité précoce (34%), la majorité tardive (34%), les retardataires (16%).

Comme l'indique le graphique (voir figure 2), le premier marché et le pionnier des assistants vocaux est celui des États-Unis. Dans les autres pays, tels que le Royaume-Uni, l'Allemagne, le Canada, le Japon et l'Australie, l'adoption était proche en 2018, voire sur le point de passer dans la tranche « Early Majority¹⁰ », qui représente 34% de la population. La particularité de cette courbe est qu'elle comprend un point mort entre les phases d'adoption nommées « Early Adopters¹¹ » et « Early majority ». Il s'agit d'un passage primordial où l'innovation quitte son marché de niche pour pénétrer son marché de masse (Thurier, 2020).

En effet, Geoffrey Moore¹² a revu le modèle proposé par Rogers, en 1991, et y a ajouté une zone entre ces deux phases, appelée le gouffre. En d'autres mots, il s'agit d'un point mort où les technologies peuvent disparaître. Il se peut donc qu'au cours de ce cycle, l'intérêt des consommateurs s'estompe et que l'adoption stagne à ce stade. Selon cette interprétation, on comprend que les assistants vocaux sont utilisés couramment par les consommateurs. Les assistants vocaux devraient ainsi se trouver à l'abri de ce gouffre s'ils suivent la tendance initiée par les USA (Ben-Yaacov et al., 2001).

⁸ Enceinte connectée, voir glossaire n°17.

⁹ Everett Rogers est un sociologue et statisticien américain.

¹⁰ En français : « Majorité précoce »

¹¹ En français : « les premiers adeptes »

¹² Geoffrey Moore est un écrivain et consultant hi-tech de la Silicon Valley

Adoption de la technologie de l'assistant vocal par pays

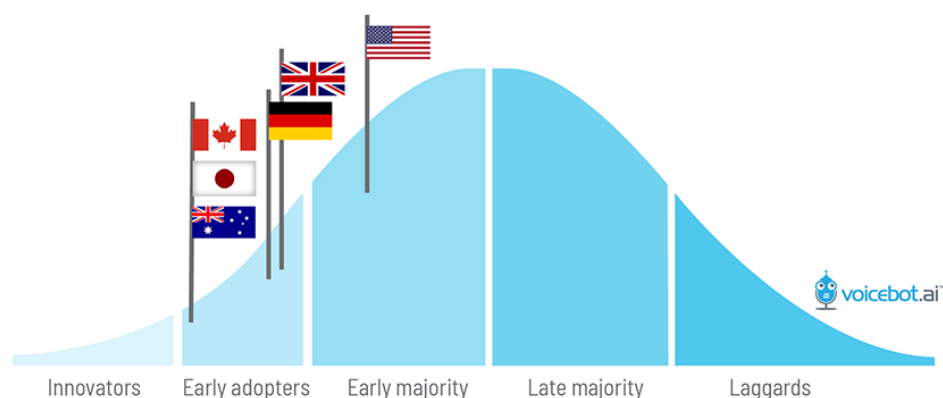


Figure 2: Adoption de l'assistant vocal par pays sur la courbe d'adoption de l'innovation de Rogers
(Source : Voicebot.ai, 2018)

En 2019, Roland berger, cabinet de conseil en stratégie, et VISEO, société de conseil et services numériques, se sont associés à dessein d'analyser le marché des assistants vocaux. En France, l'entrée sur le marché a eu lieu en 2012 avec la version française de Siri. D'après leur étude, les assistants vocaux semblaient déjà adoptés par les français 7 ans plus tard. En effet, en 2019, le nombre d'utilisateurs d'assistants vocaux de l'Hexagone variait entre 16 et 20 millions, en fonction des différents dispositifs utilisés.

Statista prévoit que le nombre d'assistants vocaux¹⁵ utilisés doublera de 2019 à 2024, au niveau mondial. En effet, dans leur étude publiée en 2020, ils présentaient les chiffres de 2019 : à l'époque, on comptabilisait 3,25 milliards d'assistants vocaux dans le monde. D'après les prévisions, 8,4 milliards d'assistants vocaux seraient utilisés en 2024 (voir figure 3). Ce chiffre n'est pas anodin car il est censé dépasser le nombre d'habitants compris dans la population mondiale (Dosquet et al., 2020, p.25).

¹⁵ Statista considère les assistants vocaux comme l'association d'un utilisateur individuel avec un assistant et un appareil particulier. Cela signifie que certains utilisateurs auront plusieurs assistants en service alors que d'autres n'en auront aucun. Il inclut également les assistants destinés aux particuliers et aux entreprises.

Évolution du nombre d'assistants vocaux

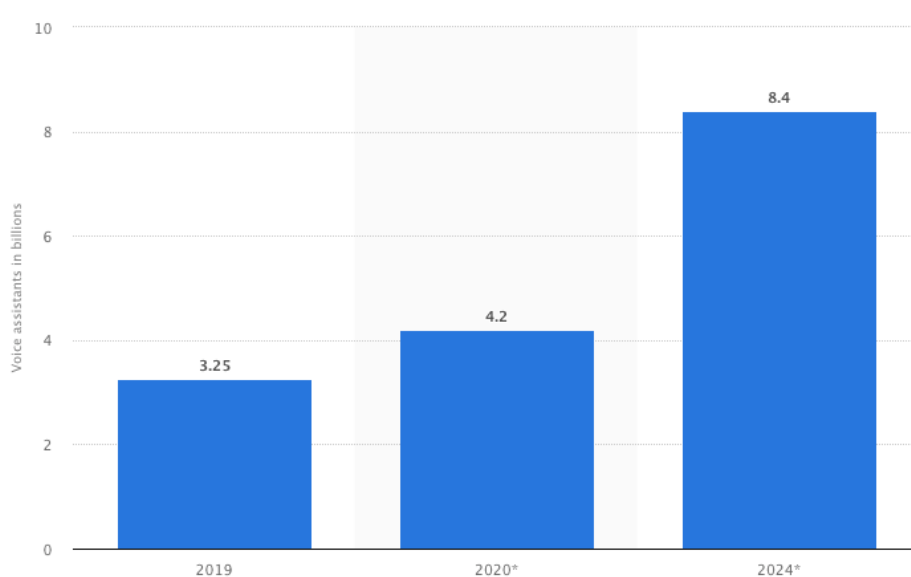


Figure 3: Nombre d'assistants vocaux numériques dans le monde de 2019 à 2024 (en milliards)
(Source : Statista, 2021)

D'après une étude réalisée par Wavemaker¹⁶ en 2020, sur un échantillon représentatif de 1000 Belges de 18 à 59 ans, le pourcentage de citoyens utilisant des assistants vocaux était de 30%. Ce pourcentage était réparti de manière égale dans le nord et le sud du pays. Au niveau des genres, il n'y avait pas de différence majeure. Par contre, l'étude met en évidence des chiffres davantage significatifs dans la tranche d'âge 18-34. Cette étude pointe également que le dispositif majoritairement utilisé, à 84%, est le smartphone, suivi par l'enceinte connectée.

2.1.3 Besoin à l'égard des nouvelles technologies

Lorsque l'on s'intéresse à une nouvelle technologie, il convient d'analyser comment se caractérise la demande à son égard. La demande peut être qualifiée de différentes manières. Les demandes les plus fréquentes sont, soit existantes, soit latentes. La première implique qu'il y a une demande déjà présente de la part des consommateurs pour un produit particulier. Quant à la seconde, celle-ci indique une demande que le consommateur réalisera plus tard. Cela signifie qu'il n'y a pas de demande pour le produit avant qu'il soit introduit sur le marché (Aunege, s.d. ; Bhasin, 2020).

¹⁶ Wavemaker est une agence de conseils en activation technologique.

La demande latente est la plus adéquate pour les assistants vocaux. En effet, les fonctionnalités qu'offre un assistant vocal seront attractives pour les utilisateurs, étant donné qu'il s'agira d'une technologie nouvelle à leurs yeux. Ils ne seront cependant pas conscients des besoins que l'assistant vocal peut satisfaire avant même d'avoir l'opportunité d'en posséder un (Bhasin, 2020).

Ensuite, il est également important d'identifier la source de l'innovation, qui est indirectement liée à la demande. Une nouvelle technologie peut voir le jour, soit en vue de créer un nouveau besoin, soit afin de répondre à un certain besoin. Le premier cas porte le nom de « technology push » et implique la commercialisation d'inventions ou de découvertes technologiques abouties. Le second cas, quant à lui, nommé « market pull », implique d'identifier les besoins du marché mieux que les autres (Gailly, 2020).

Dans le cas des assistants vocaux, il est question de technologie push. À l'instar du phonographe¹⁸, inventé par Edison, cette technologie offre de nouvelles connaissances. La technique « push » suggère de repérer une solution. Tandis que la technique « pull » indique de repérer un besoin (Gailly, 2020).

2.1.4 Adoption des assistants vocaux

Les assistants vocaux détiennent un avenir prometteur. Les possibilités d'utilisations futures de cette technologie sont nombreuses, allant des voitures connectées à la traduction en passant par la domotique¹⁹ mais également la compagnie et l'assistance aux personnes âgées (Hoy, 2018).

Afin d'établir des prévisions quant à l'adoption de l'assistant vocal dans les prochaines années, nous allons nous intéresser à la courbe de Gartner. Chaque année, la société de recherche et de conseil choisi une sélection de technologies émergentes afin d'analyser les attentes des consommateurs envers celles-ci. Également appelée « Hype Cycle²⁰ », la courbe passe par les cinq étapes qu'une nouvelle technologie est censée suivre successivement. Ces étapes représentent le cycle de vie de cette technologie. Cette courbe estime également le temps nécessaire pour que la technologie atteigne le plateau de productivité. On retrouve le niveau

¹⁸ Phonographe, voir glossaire n°35.

¹⁹ Domotique, voir glossaire n°15.

²⁰ En français : « Cycle de l'effervescence » ou « cycle de l'innovation »

d'attente, sur l'axe vertical et le temps, sur l'axe horizontal. Cette courbe permet ainsi de situer les innovations par rapport à leur niveau de maturité dans le temps (Cohen, 2021).

Afin de suivre l'évolution de l'assistant vocal, nous avons sélectionné les graphiques²¹ incluant l'assistant virtuel, établis de 2017 à 2019 par Gartner. À partir de ceux-ci, nous avons reconstruit un graphique reprenant uniquement l'assistant virtuel afin de mettre en exergue l'historique de l'évolution de la technologie à travers le temps (voir figure 4). La raison pour laquelle notre analyse s'arrête au graphique de 2019 est que Gartner ne sélectionne pas les mêmes technologies d'année en année. La majorité des technologies retenues sur la courbe de 2020 ne figuraient pas sur les courbes précédentes. L'assistant virtuel n'en fait d'ailleurs plus partie.

Évolution de l'état de maturité de l'assistant virtuel

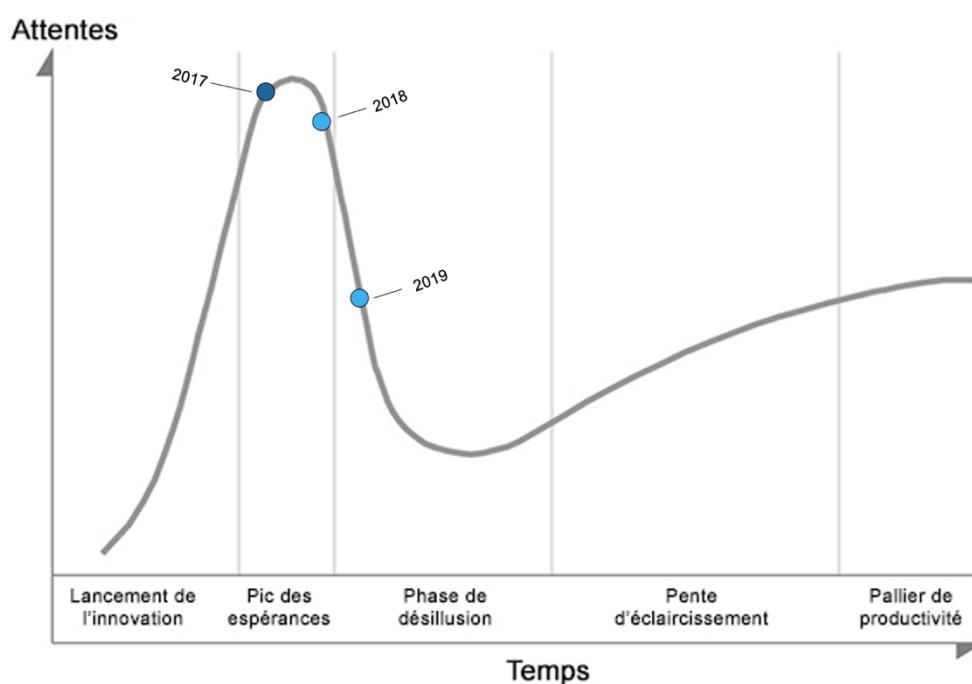


Figure 4: Évolution de l'assistant virtuel sur la courbe de Gartner, de 2017 à 2019

Au vu de leur position sur le graphique de 2017, les principales catégories de l'IA, dont les assistants virtuels, avaient atteint la deuxième phase du cycle, à savoir le pic des attentes excessives. À ce stade, cela signifie que, compte tenu de l'intérêt du marché, une nouvelle ère

²¹ Voir courbes de Gartner incluant l'assistant vocal en annexe n°2

de l'informatique s'annonce. En 2017, alors que certains cas d'usage se profilaient, on constatait que la majorité des acheteurs patientaient (Kinsella, 2017).

De 2017 à 2018, l'assistant virtuel a redescendu la pente de la courbe, tout en restant dans la deuxième phase. Placer une technologie au sommet des attentes suggère que la prochaine étape, nommée la désillusion, est proche.

Sans surprise, l'année suivante, l'assistant virtuel est entré dans la troisième étape du cycle de vie. En effet, la courbe de Gartner publiée en 2019 montre que l'assistant virtuel est placé dans la phase de désillusion. En d'autres termes, l'engouement autour de l'assistant virtuel cède place au désintérêt. Les acheteurs sont désabusés par le rythme des progrès et par les limitations de cette technologie (Kinsella, 2017). Lorsqu'elle se trouve à cette étape, il se peut que la technologie subisse une baisse de financement. L'IA a par ailleurs déjà fait les frais de pertes de financement à trois reprises dans son histoire. Ces périodes ont d'ailleurs été appelées « AI winter » (voir section 2.2.3.3). D'après Peter Cahill (2017) un énième « AI winter » n'est pas à craindre. L'IA, de manière générale n'est pas utilisée dans tout son potentiel d'application. De plus, au vu du nombre de cas d'utilisation émergeant dans les entreprises chaque mois, cela ne semble pas probable.

D'après le graphique de 2019, l'assistant virtuel est supposé atteindre le plateau de productivité d'ici 2 à 5 ans. Être sur le plateau signifie que l'innovation a fait ses preuves et que l'adoption est devenue « mainstream²² ». Si la technologie atteint la cinquième et dernière étape de la courbe, cela signifie que les risques auront pu être réduits au maximum. On pourra alors dire que l'assistant virtuel est un produit mature (Tomaz, 2019).

En janvier dernier, Gartner a sorti un autre graphique incluant l'assistant virtuel, avancé cette fois. Le graphique en question est un radar d'impact des tendances. Bien que les assistants virtuels soient sur le marché depuis un certain temps, les cas d'usage sont restreints et l'expérience peut être limitée, voire médiocre. Les assistants virtuels avancés²³, eux, occupent une place importante sur ce radar et sont repris dans la catégorie « Interfaces et expériences ». Cette catégorie inclut les technologies qui convertissent notre façon d'interagir avec le monde.

²² En français : « adoption généralisée »

²³ Les assistants virtuels avancés sont un niveau supérieur d'assistant virtuel, qui ciblent une multitude de tâches et de fonctions.

Positionné au premier rang du radar, l'assistant virtuel avancé est supposé être adopté par la grande majorité d'ici un à trois ans. Selon Gartner, l'impact d'une technologie peut être qualifiée sur une échelle de faible à très élevée. L'assistant virtuel avancé a un potentiel d'impact élevé car il peut être utilisé dans un très grand nombre de domaines (voir figure 5) (Nguyen, 2021).

Prévisions sur l'impact des assistants virtuels avancés sur le marché

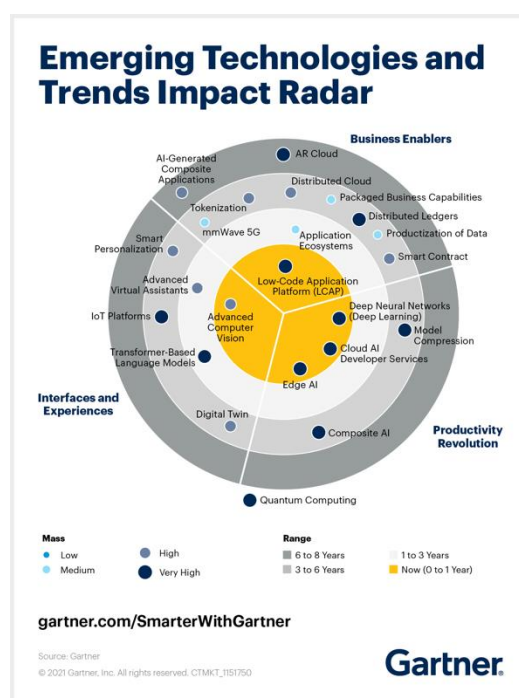


Figure 5 : Radar d'impact des technologies émergentes
(Source : Gartner, 2021)

Finalement, nous allons analyser l'évolution des autres technologies pour pouvoir mettre ces chiffres présentés ci-dessus en perspective. En 2018, Activate, une société de consultance en technologie, estimait que les enceintes connectées avaient un grand potentiel d'évolution sur le marché américain. En effet, en comparaison à tout autre appareil, les enceintes connectées prévoient d'être adoptées plus rapidement. Alors qu'il a fallu 5 années aux enceintes connectées pour être adoptées par la moitié de la population des États-Unis, il a fallu 5 ans de plus à la TV et 15 ans de plus à la radio pour atteindre le même pourcentage. En conclusion, l'évolution de l'assistant vocal est supposée être exponentielle par rapport à l'évolution de n'importe quelle technologie précédente (voir figure 6) (Dosquet et al., 2020, p.24).

Évolution de l'adoption des technologies dans le temps sur le marché américain

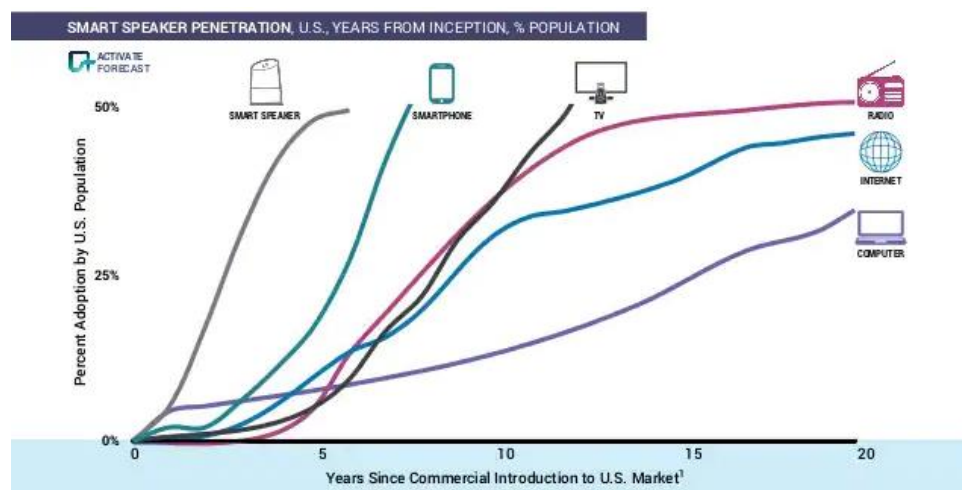


Figure 6 : Comparaison des taux d'adoption de plusieurs technologies aux USA
(Source : Activate, 2018)

Après avoir analysé les tendances sur le marché des assistants et ce dans différentes régions du monde, nous serons à même d'appréhender en profondeur les deux concepts majeurs de notre recherche, qui sont l'intelligence artificielle et l'assistant vocal, ce dernier usant des technologies de l'IA.

2.2 L'intelligence artificielle

De nos jours, l'intelligence artificielle (IA) est un sujet qui soulève beaucoup de questions et ce dans beaucoup de domaines. Comme l'émet Le Cun (2019) dans son ouvrage, « *l'intelligence artificielle est en train de coloniser tous les secteurs de l'économie, de la communication, de la santé, et des transports avec les voitures autonomes... Beaucoup d'observateurs ne parlent plus d'une évolution technologique, mais d'une révolution.* » L'intelligence artificielle ne cesse de s'immiscer de manière significative dans davantage de domaines.

Dans cette partie, nous fournirons premièrement une définition complète de ce que représente l'IA, comprenant les différents types d'IA existants ainsi que leurs caractéristiques. Ensuite, nous allons mettre en exergue les moments clés de l'histoire des découvertes liées à l'intelligence artificielle.

Tout au long de ce document, l'acronyme IA sera utilisé pour définir l'intelligence artificielle. Cela nous permettra d'éviter toute confusion par la suite.

2.2.1 Définition

Selon la Commission Européenne, l'intelligence artificielle est décrite comme telle : « *L'IA désigne les systèmes qui font preuve d'un comportement intelligent en analysant leur environnement et en prenant des mesures - avec un certain degré d'autonomie - pour atteindre des objectifs spécifiques. Les systèmes basés sur l'IA peuvent être purement logiciels et agir dans le monde virtuel (par exemple, assistants vocaux, logiciels d'analyse d'images, moteurs de recherche, systèmes de reconnaissance de la parole et des visages) ou l'IA peut être intégrée dans des dispositifs matériels (par exemple, robots avancés, voitures autonomes, drones ou applications de l'internet des objets).* » (AI 4 Belgium, 2019, p.6).

Cette définition désigne, au sens large, les systèmes émanant du domaine de la science-fiction. Elle englobe, indistinctement, les IA dites « fortes » ainsi que les IA qualifiées de « faibles » ou « étroites ». Ces dernières sont opérationnelles et sont donc en mesure d'exécuter des tâches particulières telles que, par exemple, la reconnaissance faciale et vocale ou la conduite de véhicule (Conseil de l'Europe, s.d.). En d'autres mots, cette catégorie d'IA se consacre à un

domaine précis et se calque sur la cognition humaine, bien qu'elle ne soit pas dotée de sa conscience. L'IA faible permet de traiter des bases de données pour en sortir des informations exploitables. Pour ce faire, elle exploite des modèles et émet des prédictions. Ce type d'IA rencontre néanmoins certaines limites. Outre ses capacités limitées, il est possible qu'elle cause des dommages lorsque le système est défaillant. Elle fait aussi l'objet de préoccupations quant à la perte d'emploi croissante générée par l'automatisation de certaines tâches (Frankenfield, 2021).

Tandis que l'IA forte, également qualifiée de véritable ou nommée intelligence générale artificielle (IGA), est une forme théorique d'IA, qui elle égalise l'intelligence humaine. L'IA sous cette forme n'a pas encore vu le jour et est considérée comme le niveau suivant de l'IA. D'après certains experts, elle pourrait être développée d'ici 2030 ou 2045. D'autres, moins optimistes, estiment qu'elle sera plutôt à même de voir le jour au cours du siècle prochain. Les caractéristiques intrinsèques à l'IA sont la capacité de raisonner, de résoudre des énigmes, de juger, de planifier, d'apprendre et même de communiquer. Inversement à l'IA faible, celle-ci doit être munie d'une conscience. Les avis sont toutefois assez divergents en la matière. Certains théoriciens émettent l'hypothèse que des machines qui seront dotées d'une IA forte pourront grandir en passant d'un esprit puéril à un esprit adulte, en suivant le processus de développement de l'être humain (Frankenfield, 2020).

Longtemps avant que l'IA ne soit divisée entre ces deux classifications, elle était qualifiée de « Good Old-Fashioned AI » de l'acronyme « GOFAI²⁴ », traduit « bonne intelligence artificielle à l'ancienne ». Le terme, attribué en 1985 par John Haugeland, désigne les IA qui sont programmées pour fournir une réponse à une requête. Ces dernières ne sont pas en mesure d'apprendre ou de modifier elles-mêmes leur programmation. Ce type d'IA a d'ailleurs été le plus répandu en termes d'utilisation ces 50 dernières années. Bien que cette approche « GOFAI » de l'IA symbolique soit l'aînée, cette méthode classique est encore largement utilisée (Future of Tech, s.d.). En d'autres mots, l'IA symbolique fonctionne selon un système de règles conçues et intégrées aux machines par les humains. Ces règles ont pour but d'amener la machine à la prise de décision autonome. Ainsi, l'IA répond selon les informations reçues (Verani, 2020).

²⁴ GOFAI, voir glossaire n°21.

2.2.2 Les algorithmes

Dans un article, la Commission Nationale de l'Informatique et des Libertés²⁵ (CNIL, s.d.) définit un algorithme comme étant « la description d'une suite d'étapes permettant d'obtenir un résultat à partir d'éléments fournis en entrée ». Cette définition semble vulgariser le terme afin de faciliter sa compréhension. La commission illustre d'ailleurs l'algorithme avec la métaphore de la recette de cuisine qui donne une série d'instructions afin d'obtenir un plat, un résultat.

L'intuition de l'algorithme a vu le jour grâce à Alan Turing. Il s'agit d'un mathématicien britannique. C'est d'abord en 1928 et sous l'impulsion du défi lancé par le mathématicien David Hilbert à ses collègues, que Alan s'est lancé dans ses recherches autour de l'informatique. Ce défi, nommé « problème de la décision » pose la question de l'existence d'une machine qui puisse déterminer si un énoncé mathématique est correct ou erroné. Hilbert était alors persuadé que la réponse était positive. Tandis que Turing, convaincu du contraire, souhaitait réfuter cette hypothèse. Il commence alors ses recherches au King's College de l'université de Cambridge. (CNRS, 2012, p.9).

En 1936, le Britannique confirme son intuition à travers son article publié dans les Annales de la Société de mathématiques de Londres. Afin de démontrer son raisonnement, il a conçu, de manière imaginaire, une machine, qui lui a permis d'expliquer un calcul complexe sous forme d'une séquence d'opérations simples. En d'autres mots, il a mis en lumière le fait qu'un calcul, qui peut être résolu par un humain, peut aussi l'être grâce à une machine. Sa démonstration n'a pas nécessité la construction de cette machine, seul son raisonnement lui a suffi. Malgré cette réponse, il n'avait pas encore toutes les cartes en main pour résoudre le problème de décision initié par David Hilbert. Jusqu'alors, la machine imaginée ne pouvait répondre qu'à un seul genre de calcul pour lequel elle aurait été conditionnée (CNRS, 2012, p.9).

Dès lors, le mathématicien prévoit une machine universelle capable d'imiter n'importe quelle autre machine. En ce qui concerne son fonctionnement, il faut lui transmettre le programme qu'elle est supposée réaliser sous format codé. Ainsi, le concept d'ordinateur est apparu et persiste encore à l'heure actuelle. Il est question d'une machine à laquelle on fournit un programme, une forme codée (aujourd'hui appelée algorithme) ainsi que les données à traiter.

²⁵ CNIL, voir glossaire n°11.

La machine qu'il a conçue, qualifiée de généraliste, lui a alors permis de s'atteler au problème de Hilbert (CNRS, 2012, p.9).

En réponse au problème de Hilbert, il prouve, premièrement, qu'il n'existe pas d'algorithme qui permette de définir si un calcul peut se conclure à un moment donné. Il n'y a donc pas de solution existante au problème de Hilbert (CNRS, 2012, p.9).

2.2.3 Historique et évolution

Dans cette section nous retracerons les points clés de l'évolution de l'IA, de sa création aux dernières avancées en la matière. Le texte qui suit ne prétend pas être exhaustif quant à la liste des dates qui ont marqué l'histoire de l'IA. Nous mettrons également en lumière les principaux acteurs.

2.2.3.1 Naissance de l'intelligence artificielle

La discipline qu'est l'intelligence artificielle est jeune d'une soixante d'années. Les pères fondateurs de la technologie qui sous-tend l'IA, ne sont autres que Alan Turing et John Von Neumann. À l'époque, il n'était pas encore question d'intelligence qu'une machine pourrait posséder. En effet, Turing a d'abord dû formaliser la notion d'algorithme (Guerraoui & Hoang, 2020).

En revanche, dans son article intitulé « Computing Machinery and Intelligence », paru en 1950, Alan Turing émet la première hypothèse d'une éventuelle intelligence émanant d'une machine. En effet, dans le prélude de ce célèbre article, il s'interroge sur la question suivante : « les machines peuvent-elles penser ? ». Étant donné que la définition du sens des mots « machine » et « penser » lui semblait très compliquée, il a décidé de remplacer cette question par un procédé expérimental moins ambigu (Marquis et al., 2009, p.1).

Il présente un dispositif capable de démontrer ou réfuter si une machine peut penser et par conséquent, faire preuve d'intelligence. Ce test est à l'origine du jeu de l'imitation, plus connu aujourd'hui sous le nom de « Test de Turing », a pour but de faire passer la machine pour un humain. Trois personnes sont impliquées dans ce jeu : un homme (A), une femme (B) ainsi

qu'un interrogateur (C), quel que soit son genre. Le but de l'interrogateur (C), isolé dans une pièce, est de déterminer, au terme du jeu, le sexe de son interlocuteur. Autrement dit, il doit identifier si les humains interrogés, connus sous le nom de X et Y, sont A ou B. Tandis que les personnes interrogées (A et B) ont pour consigne d'essayer de tromper l'interrogateur dans leur réponse, de sorte à fausser l'identification. Pour éviter que l'interrogateur soit influencé par le ton de la voix, les réponses fournies par X et Y sont écrites ou dactylographiées ou, au mieux, un téléimprimeur communique entre les deux pièces. Il existe une alternative : les réponses aux questions peuvent être répétées à l'interrogateur par l'intermédiaire d'une personne tierce (Bertrand, 1987 ; CNRS, 2012).

Ensuite, vient le tour de la machine d'endosser le rôle de l'interlocuteur de genre masculin (A). La question investiguée devient alors « Le juge parviendra-t-il plus facilement à distinguer l'ordinateur de l'humain, que l'homme de la femme ? » et remplace la question initiale « Les machines peuvent-elle penser ? » (Bertrand, 1987 ; CNRS, 2012).

Le test est considéré comme réussi quand, à circonstances égales, la machine se trompe plus souvent pour distinguer une femme d'un homme. Si le juge n'est pas en mesure de distinguer si son interlocuteur est une machine ou un être humain, alors la machine est considérée comme une entité intelligente (Epstein et al., 2009).

Dans un but empirique, le test de Turing a tenté de faire rivaliser une machine avec l'humain de par ses performances dans différents domaines supposés requérir de l'intelligence. Bien que la question à l'origine du test eût un caractère philosophique, elle a ouvert la piste de l'IA et a guidé plusieurs générations de chercheurs dans leurs accomplissements (Geffroy, 2012, p.13).

Officiellement, le terme « intelligence artificielle » a vu le jour en 1956 à la Conférence de Dartmouth organisée au Dartmouth College par Marvin Minsky et John McCarthy, qui a lui-même proposé l'appellation. C'est au sortir de la seconde guerre mondiale que cet atelier scientifique a eu lieu (Le Cun, 2019, p.28).

La même année, le premier logiciel d'IA, Logic Theorist (le « théoricien logique ») est créé par deux chercheurs américains de l'université de Carnegie-Mellon à Pittsburgh. Allen Newell et

Simon Herbert présentent un logiciel capable de démontrer des théorèmes mathématiques simples (Le Cun, 2019, p.29).

2.2.3.2 Premières inventions d'intelligence artificielle

À partir des années 60, les recherches sur l'IA ont pu être financées par le Département de la Défense en Amérique (Blanchot, 2018). C'est aussi à ce moment-là que l'innovation en la matière a connu une grande croissance. De nouveaux langages de programmations, des robots ainsi que les études de recherches ont commencé à devenir plus populaires. Cela a eu un grand impact à l'égard de l'importance accordée à l'IA dans la seconde moitié du 20^{ème} siècle (Reynoso, 2019).

2.2.3.3 Hiver de l'Intelligence Artificielle (AI Winter)

L'histoire du développement de l'IA a connu une période de crise en 1974. On lui a d'ailleurs donné le nom de « AI Winter ». En effet, durant quelques années voire décennies, un manque d'intérêt pour la recherche sur l'IA s'est fait ressentir et a conduit à une perte de financement. Pendant ce temps, la recherche et le développement se sont concentrés sur d'autres domaines. L'élément déclencheur de cette perte de financement aurait été la déception des financeurs quant aux résultats négatifs des machines soumises au test de Turing (Frankenfield, 2020).

Lorsque des sources de financement s'amenuisent, les recherches sont poursuivies par les universitaires les plus dévoués, laissant place à un rythme de l'innovation lent ou, pour le moins, ralenti (Frankenfield, 2020).

2.2.3.4 La renaissance de l'Intelligence Artificielle

Peu après que cette période sombre soit arrivée à son terme, en 1997, le logiciel « Deep Blue » de IBM²⁶ (International Business Machines Corporation) a battu le joueur d'échecs, Garry Kasparov. Ensuite, dès le début des années 2000, l'IA a fait son apparition dans bon nombre de films de science-fiction où l'on mettait en scène des scénarios tantôt réalistes, tantôt irréalistes (Blanchot, 2018).

²⁶ IBM, voir glossaire n°24.

Depuis 2010, on peut dire que les technologies qui relèvent de l'IA font partie intégrante de notre quotidien. En effet, c'est en utilisant nos smartphones dotés de reconnaissance faciale et/ou vocale que nous mettons à l'œuvre des fonctions d'IA, qui sont mêmes considérées comme acquises par le plus grand nombre (Reynoso, 2019).

2.2.3.5 Les algorithmes d'apprentissage

C'est grâce aux algorithmes d'apprentissages automatiques qu'en 2010 l'IA a connu un nouvel essor. L'IA doit cette avancée à, premièrement, l'accès à un volume massif de données, et deuxièmement, à l'efficacité des processeurs de cartes graphiques, qui permet d'accélérer le calcul des algorithmes d'apprentissage. Ces deux facteurs ont initié un nouvel engouement de la part des chercheurs et de l'industrie informatique (Conseil de l'Europe, s.d.).

2.2.3.6 L'Intelligence Artificielle non symbolique

Auparavant, seules les approches symboliques étaient d'usage. Elles consistaient à se servir de représentations symboliques du monde extérieur afin de reproduire ces opérations. Tandis que les approches non symboliques se servent de l'analyse statistique dans un but d'imitation de la capacité d'apprentissage telle qu'elle existe dans un cerveau humain. Les approches non symboliques sont grandement utilisées dans l'apprentissage automatique²⁷ ainsi que dans l'apprentissage profond²⁸, qui sont deux domaines qui sous-tendent l'IA (Future of Tech, s.d.).

2.2.3.6.1 L'apprentissage-machine/ automatique (Machine learning)

Ces nouvelles technologies dont fait partie l'IA relèvent de l'apprentissage-machine²⁹. Pour pouvoir construire une machine dotée d'une intelligence comparable à celle d'un être humain, il faut que celle-ci soit en mesure d'apprendre. Le cerveau humain est composé d'un réseau de neurones interconnectés. Ces connexions entre les neurones sont appelées synapses.

²⁷ Apprentissage automatique ou apprentissage-machine, voir glossaire n°4.

²⁸ Apprentissage profond, voir glossaire n°5.

²⁹ Apprentissage-machine ou apprentissage automatique, voir glossaire n°4.

L'apprentissage se réalise sur base de création, suppression ou encore par modification de ces synapses (Le Cun, 2019, p.20).

Le principe de l'apprentissage-machine, appelé également apprentissage automatique, consiste en la création de réseau de neurones artificiels. La première étape d'apprentissage du « machine learning³⁰ » est un entraînement durant lequel la machine apprend à effectuer une tâche au fur et à mesure. Deuxièmement, une fois l'apprentissage accompli, la machine le met en pratique (Le Cun, 2019, p.20).

Concrètement, l'apprentissage consiste à présenter à la machine des milliers d'images sur lesquelles on peut clairement identifier un objet ou une personne, par exemple un avion ou une voiture. À chaque image fournie, la machine attribue une réponse en sortie. Lorsque la réponse est erronée, on effectue un ajustement des paramètres internes de la machine afin que sa sortie corresponde à la réponse attendue. Une fois les images traitées, la machine est capable de reconnaître l'objet sur l'image présentée, qu'elle soit déjà connue de l'appareil ou non. Il s'agit de la capacité de généralisation (Le Cun, 2019, p.20).

Il est possible de classer les algorithmes d'apprentissage automatique en deux catégories : supervisés et non supervisés. Dans les algorithmes supervisés, les données d'apprentissage sont étiquetées avec les réponses. Tandis qu'avec les algorithmes non supervisés, les étiquettes éventuelles ne sont pas montrées à l'algorithme d'apprentissage (Heller, 2020).

Le fonctionnement de l'apprentissage supervisé est défini comme tel par Georgios Petropoulos (2017) : *« le réseau neuronal reçoit des exemples d'entrées et de sorties souhaitées correspondantes. Il apprend ensuite à faire correspondre précisément les entrées aux sorties en ajustant les poids et les seuils d'activation de ses connexions neuronales »*.

Il enchaîne dans son article sur le sujet de l'apprentissage non supervisé, qu'il décrit de la sorte : *« le réseau neuronal reçoit des exemples d'entrées, puis il doit reconnaître les caractéristiques, les modèles et la structure de ces entrées sans aucune indication spécifique. Ce type d'apprentissage peut être utilisé pour regrouper les données d'entrée en classes sur la base de leurs propriétés statistiques. Il est particulièrement utile pour trouver des choses dont*

³⁰ Machine learning, voir glossaire n°29.

vous ne connaissez pas la forme, comme des modèles non encore reconnus dans un grand ensemble de données » (Petropoulos, 2017).

2.2.3.6.2 L'apprentissage profond (Deep learning)

L'apprentissage profond est une sous-catégorie de l'apprentissage automatique, tout comme ce dernier est une sous-catégorie de l'IA (Voir figure 7).

Sous-catégories de l'intelligence artificielle

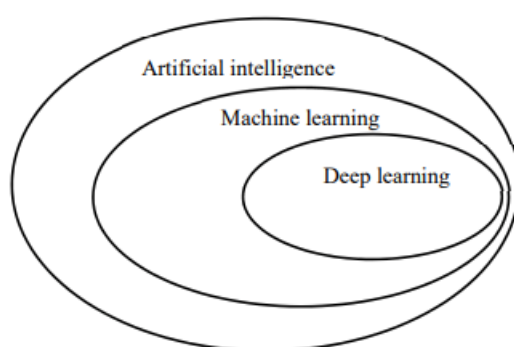


Figure 7: Diagramme de Venn de l'intelligence artificielle
(Source : Mellit & al., 2020)

Cette sous-catégorie de l'IA suit le même principe que l'apprentissage automatique. Elle est un moyen de fournir des connaissances aux ordinateurs via des données et observations. La particularité de l'apprentissage-machine est qu'il consiste à apprendre aux machines comment se former, à partir de leurs expériences, tel que les humains le font (Fellous, 2021).

À la différence de l'apprentissage automatique, les algorithmes d'apprentissage profond ne sont pas restreints en termes de capacité d'apprentissage. Au plus les systèmes reçoivent de l'information et des données à traiter, au plus ils sont supposés augmenter leur performance (Datascientest, 2021).

Le traitement naturel du langage (ci-après TALN³¹) s'appuie sur les techniques de Deep Learning³². Le TALN étudie le langage humain à travers nos comportements (IA School, 2021).

³¹ TALN, voir glossaire n°40.

³² Deep Learning, voir glossaire n°13.

Nous allons expliquer plus en détail dans les prochaines sections en quoi cette technologie de l'IA est utile au fonctionnement de l'assistant vocal (voir section 2.3.3.3).

2.2.3.7 Big data

Une fois que les chercheurs ont eu trouvé des procédés efficaces afin de mettre en œuvre l'apprentissage machine et ensuite l'apprentissage profond (voir sections 2.2.3.6.1 et 2.2.3.6.2), il a été possible de fournir aux machines un nombre plus conséquent de données. À ce jour, la problématique concernée par l'IA n'étant plus de posséder les systèmes performants mais bien d'avoir les données à traiter (Blanchot, 2018). Arrivées à un certain point, les données n'étaient plus suffisantes pour être en mesure de prédire avec exactitude les résultats (Edureka, 2018).

C'est alors que l'industrie a commencé à aller de l'avant. Google, par exemple, en possession de données, est devenu l'un des pionniers de l'IA. En 2013, Facebook a décidé de lancer une activité dans la recherche et le développement. Ainsi, le centre Facebook Artificial Intelligence Research (FAIR), dirigé par Yann Le Cun, a vu le jour. En d'autres termes, il s'agit de sociétés qui possèdent d'énormes volumes de données (Blanchot, 2018).

On parle désormais de l'ère du big data³³. Cette notion est toutefois à considérer avec précaution. En effet, il est important de nuancer big data de quality data³⁴. Originellement, le big data sous-entend le volume important de données. On lui attribue toutefois d'autres caractéristiques déterminées par les 3V's, à savoir le Volume, la Vitesse et la Variété. Le big data est supposé amener une plus-value aux entreprises. Cependant, pour que ça soit une réelle plus-value, il ne faut pas négliger la qualité de ces données, qui représente un réel challenge. C'est pourquoi, plus tard, deux caractéristiques additionnelles se sont greffées. Les 5V's fournissent ainsi les caractéristiques complètes, auxquelles on a ajouté la Véracité ainsi que la Valeur. Ces dimensions servent à décrire l'intégrité et la qualité des données (Janev et al., 2016).

³³ Big Data, voir glossaire n°7.

³⁴ Quality Data, voir glossaire n°36.

2.2.3.8 Le futur de l'intelligence artificielle

Les experts estiment que de nouvelles découvertes en recherches fondamentales sont nécessaires afin d'amener l'IA à l'imitation de la cognition humaine, c'est-à-dire une IA de type « forte » (voir section 2.2.1). En effet, une évolution des performances technologiques en matière d'apprentissage automatique ne permettrait pas une telle avancée (Conseil de l'Europe, s.d.).

À l'heure actuelle, l'IA représente un enjeu sociétal. Alors que nous ne sommes qu'aux prémices de l'intégration de ces nouvelles technologies, les recherches scientifiques laissent place aux débats publics. En effet, les IA, de par leurs performances dans un grand nombre de domaines, nous amènent à un réel questionnement quant au rôle professionnel de l'être humain dans les décennies à venir (Blanchot, 2018). D'après Frankenfield (2020), l'IA ne serait pas à l'abri de connaître un second « hiver », tel que celui vécu en 1974 (voir section 2.3.3). Cependant, tant qu'elle sera considérée par les entreprises comme une source de profit ou une économie potentielle, l'IA devrait éviter une telle perte de financement.

2.3 Les assistants vocaux

Ces dernières années, nous avons connu une révolution dans la manière d'interagir avec les différentes interfaces, en particulier à l'aide de la voix. Il est important d'avoir décrit l'intelligence artificielle dans la section précédente car les assistants vocaux sont soutenus par l'IA. En effet, grâce au couplage avec l'IA, les interfaces textes et vocales connaissent des utilisations bien différentes.

Assistant vocal, assistant intelligent ou encore assistant personnel, les synonymes sont nombreux pour désigner l'assistant vocal. Celui-ci peut être utilisé par le biais d'une enceinte connectée, un smartphone ou une autre interface. Cette nouvelle ère des assistants vocaux est bien présente. Au fur et à mesure, les différents secteurs sont investis, tant dans le public que dans le privé.

À travers cette section, nous allons décrire le sujet principal de notre mémoire : les assistants vocaux. Nous retracerons leur développement au fil du temps. Afin de comprendre les tenants et aboutissants liés à cette technologie, nous listerons les limites liées à son utilisation. Nous nous intéresserons au marché des assistants vocaux en mettant en lumière les usages principaux ainsi que les cadres d'utilisation. Nous aborderons également le sujet de la sécurité et de la confidentialité dans le traitement des données confrontées à l'assistant vocal.

L'objectif principal de cette partie théorique au sein de la revue de littérature est de saisir le rôle déterminant que joueront les assistants vocaux dans les cas d'usages étudiés dans notre analyse (voir Chapitre 4).

2.3.1 Définition

Afin d'appréhender de manière concrète ce qu'est réellement un assistant vocal, il est opportun d'acquérir une certaine compréhension de ce terme. Le terme complet est en réalité « assistant virtuel à commande vocale », où l'assistant virtuel et l'interface vocale sont les deux éléments combinant le terme initial. Dès lors, il est nécessaire de clarifier le terme complet en mentionnant la définition proposée et expliquée par Dosquet et al. (2020) dans leur ouvrage à propos de la révolution des assistants vocaux.

Les auteurs prônent une explication d'un assistant virtuel comme étant « *un programme informatique qui s'appuie sur plusieurs technologies d'intelligence artificielle pour effectuer des tâches et/ou rendre des services à l'utilisateur. Pour cela, il interprète ses requêtes (sous forme de texte, voix ou tactile) et interagit avec lui en mode conversationnel* ». Ce système de conversation est garanti par le biais de trois différentes interfaces, à savoir le clavier, des structures non-verbales et la voix (Dosquet et al., 2020, p.14). Dans l'élaboration de ce mémoire, nous allons uniquement nous concentrer sur cette dernière.

En conséquence de cette première définition, les auteurs déduisent qu'un assistant vocal est alors « *une combinaison d'agents intelligents (dont chacun exécute une tâche spécifique), d'applications (dont les applications vocales) et de connections entre les données* » (Dosquet et al., 2020, p.14).

En effet, l'idée principale repose sur le fait que des enceintes connectées et/ou smartphones sont conçus à l'effet d'exploiter ces logiciels intelligents. Ces derniers sont constamment à l'écoute d'un mot clé qui permettra aux appareils de se réveiller et réagir à la demande de l'utilisateur. Une fois le mot clé entendu, l'appareil enregistre la voix de cet utilisateur et la transmet à un serveur spécialisé, qui la traite et la traduit comme une instruction. Selon la demande, le serveur fournit les informations adéquates à l'assistant vocal. Enfin, l'assistant va alors exécuter ces informations et diffuser la musique demandée, lire le renseignement requis ou encore effectuer des tâches à l'aide de divers services et appareils connectés (Hoy, 2018).

Il est également opportun d'apporter quelques précisions quant à l'assistant vocal dans la pratique. Ce dernier peut également être qualifié comme étant un « assistant personnel intelligent ». En revanche, la réciproque ne peut être employée (CNIL, 2020).

2.3.2 Historique et évolution

Bien que le questionnement à l'égard du traitement de la voix ne soit pas neuf, l'avènement des assistants vocaux amène une nouvelle perspective. Afin de parvenir à comprendre les enjeux engendrés par ces nouveaux dispositifs vocaux, il est nécessaire de comprendre leur origine ainsi que leur fonctionnement. C'est pourquoi nous allons retracer l'histoire des assistants

vocaux en mettant en lumière quelques dates clés. De nombreuses étapes ont été nécessaires afin de faire voir le jour aux assistants vocaux. C'est seulement à partir de 2010 qu'ils ont été introduits dans l'imaginaire collectif (CNIL, 2020). Cet historique n'a pas la prétention de lister de manière exhaustive les dates qui ont marqué l'histoire des assistants vocaux.

L'histoire des assistants vocaux remonte à plusieurs siècles, bien qu'à l'époque il n'était pas encore question d'assistants vocaux à proprement parler. En effet, il a tout d'abord fallu développer les capacités de traitement du langage naturel, d'enregistrement numérique des sons, de transcription de la parole, etc. C'est Wolfgang von Kempelen qui a créé la première machine mécanique permettant la synthèse vocale en 1769. Quelques années plus tard, en 1869, Thomas Edison, accompagné de Charles Cros, a inventé le phonographe. C'est cet appareil acoustique qui est capable de reproduire les sons enregistrés (CNIL, 2020).

Plus récemment, en 2002, Microsoft a intégré des outils de transcription automatique dans la suite Office, convertissant la parole en texte (CNIL, 2020).

En 2011, Watson d'IBM a mis sur pied le premier programme de traitement du langage naturel, champion du jeu télévisé américain Jeopardy. La même année, le premier assistant vocal a été mis à disposition du grand public sur iOS³⁷. Apple, avec son assistant vocal nommé Siri, a inauguré l'arrivée de cette technologie sur le marché. La version française est, quant à elle, apparue une année plus tard. Les GAFA³⁸ ont ensuite successivement développé leur propre dispositif. Google a lancé sur le marché Google Now, en 2012, un assistant vocal téléchargeable sous forme d'applications Android et iOS, suivi par Amazon, qui a créé Alexa et Echo en 2014. Il s'agit d'une enceinte connectée dotée d'un assistant vocal, bien connu du grand public à l'heure actuelle (CNIL, 2020).

À partir de 2018, plusieurs marques ont signé des partenariats avec Google afin d'intégrer des assistants vocaux dans leur produits et services. Des constructeurs automobiles ont créé une alliance entre Renault, Nissan et Mitsubishi et ont ainsi pu équiper leurs véhicules de Google Assistant et Android Auto. En 2020, le groupe Carrefour a annoncé son partenariat avec le

³⁷ iOS, voir glossaire n°27.

³⁸ GAFA, voir glossaire n°19.

géant du web. Le but était de permettre aux consommateurs d'effectuer leurs achats à l'aide de Google Assistant (CNIL, 2020).

Les assistants vocaux se développent toujours plus, à l'heure d'aujourd'hui, tant par les géants comme les GAFA, que par des start-ups.

2.3.3 Le fonctionnement d'un assistant vocal

Au fur et à mesure que les technologies évoluent, notre façon d'interagir avec celles-ci se développe. Lorsque les premiers assistants technologiques ont vu le jour, notre seul moyen de les utiliser était au travers du texte. À l'heure actuelle, la voix semble être plus utilisée. Les assistants vocaux sont configurés de manière à réagir à l'entente de certains mots d'éveil.

Afin de pouvoir analyser les tenants et les aboutissants des usages de l'assistant vocal, il est important de comprendre comment celui-ci fonctionne et d'en savoir plus sur le langage utilisé dans les interactions avec l'assistant vocal. Les composantes de la reconnaissance vocale sont multiples et l'assistant vocal exige la complémentarité de plusieurs technologies à sa mise en action, du moment où l'utilisateur parle au déroulement de l'action demandée (Chapuzet, 2019). Ces différentes technologies vont être expliquées chronologiquement dans ces prochains points.

Pour une meilleure perception, le schéma suivant (voir figure 8) illustre de façon simplifiée le fonctionnement d'un assistant vocal.

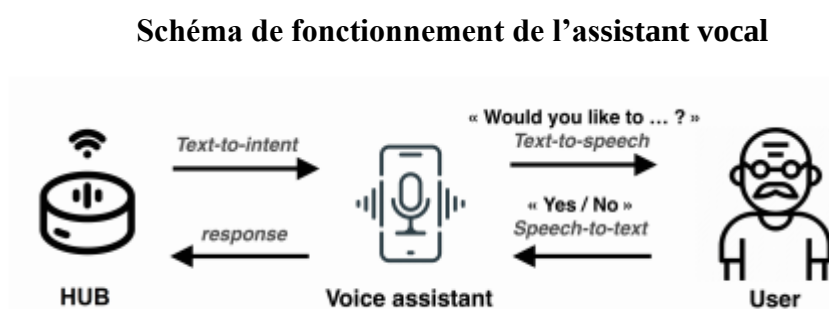


Figure 8 : Fonctionnement de l'assistant vocal, de la reconnaissance vocale à la synthèse vocale
(Source : Goffinet & Schmitz, 2021)

2.3.3.1 Reconnaissance vocale (Speech-to-text)

La première étape du processus débute lorsque l'utilisateur s'adresse à l'assistant vocal. Pour ce faire, l'individu doit avoir connaissance du mot déclencheur, plus communément appelé « Wake Word » en anglais. Cette première brique technologique a pour but de déclencher l'enregistrement de ce qui suit ce mot spécifique, entre autres par la voix. Autrement dit, il est question que l'utilisateur active le système afin qu'il puisse enregistrer la demande qui va lui être soumise. Il est à noter que ce n'est pas le seul moyen d'activer l'appareil. L'enregistrement peut également se mettre en marche une fois que la personne a appuyé sur un bouton (Chapuzet, 2019). Les assistants vocaux les plus communs sont ceux dont les enceintes connectées sont équipées. Parmi les plus connues, on retrouve Google Home et Alexa, développées par Google et Amazon, dont les mots d'éveil sont respectivement « Ok Google » et « Hey Alexa » (Kok et al., 2009).

Une fois la parole enregistrée, elle va être numérisée via le Speech-to-Text (STT³⁹). Au cours de cette étape, la voix est captée sous forme de fréquences sonores (Chapuzet, 2019). La reconnaissance vocale fonctionne à l'aide d'algorithmes par le biais de la modélisation acoustique et de la modélisation du langage, afin de reconnaître et distinguer les mots dictés. Le composant acoustique met en correspondance les unités acoustiques avec les phonèmes⁴⁰ existants. Quant au composant linguistique, ce dernier transforme ces unités acoustiques en mots, phrases et paragraphes (Amberscript, 2021 ; Kiwak 2020).

En d'autres mots, le processus établi lors de la reconnaissance vocale est alors l'identification de mots prononcés et de les convertir en un texte lisible et compréhensible (Kiwak, 2020). Dès que la reconnaissance vocale a eu lieu, c'est-à-dire lorsque les paroles de l'utilisateur ont été retranscrites en format texte, vient ensuite l'étape qui consiste à mettre en relation ce texte avec un ensemble d'intentions prédéfinies. La façon dont on s'adresse aux assistants vocaux repose sur le langage naturel (voir section 3.3.1.1), qui n'est pas censé être compris par une machine à premier abord. Les techniques de Traitement Automatique du Langage Naturel (TALN) et de Compréhension du Langage Naturel (CLN⁴³) sont donc utilisées pour rendre ce langage

³⁹ STT, voir glossaire n°38.

⁴⁰ Phonème, voir glossaire n°34.

⁴³ CNL, voir glossaire n°9.

compréhensible pour l'assistant vocal. Ces méthodes, qui facilitent la compréhension de notre langage, vont être expliquées dans les prochains points.

2.3.3.2 Le langage naturel

Pour parvenir à comprendre comment est traitée une demande soumise à un assistant vocal par la voix, il est important d'introduire la notion de langage naturel. Dans leur ouvrage, Kamath et al. (2019) décrivent la langue naturelle comme étant « *une langue qui a évolué naturellement à travers l'utilisation quotidienne par des humains au fil du temps, sans construction formelle* ». Il s'agit donc du langage emprunté par l'être humain pour s'adresser à la machine. Ce langage naturel va subir un traitement automatique ainsi que le système de compréhension.

Par après, il est également important de faire la distinction entre le TALN et la CLN. L'un permet de comprendre grammaticalement le message adressé de l'homme à la machine tandis que l'autre permet une compréhension sémantique de la demande. Cette deuxième étape reste la plus compliquée à l'heure actuelle et fait l'objet de recherche (Ioualalen, 2021). Nous allons développer le fonctionnement inhérent à ces deux disciplines dans les points suivants.

2.3.3.3 Traitement Automatique du Langage Naturel

Afin de comprendre le fonctionnement des assistants vocaux, il est important d'introduire la notion de traitement automatique du langage naturel (TALN). Jake Frankenfield (2020) explique qu'il a été conçu pour élaborer des logiciels qui génèrent et décryptent les langages naturels. Grâce à cela, un utilisateur peut converser de manière naturelle avec un ordinateur sans utiliser quelque programmation ou langage artificiel. Lorsque le TALN traite le langage écrit, on appelle cela « analyse de texte » tandis que lorsque le langage parlé est traité, on appelle cela « analyse de la parole » (Kamath et al., 2019).

Le TALN met en œuvre à la fois l'IA avec la linguistique computationnelle⁴⁴ et avec l'informatique pour parvenir à traiter les langues et les discours humains, dits naturels. Le processus suivi par le traitement automatique du langage naturel est composé de trois étapes.

⁴⁴ Linguistique computationnelle, voir glossaire n°28.

Premièrement a lieu le processus de conversion de la parole en texte. Une fois le langage naturel reçu par l'ordinateur, celui-ci use d'un modèle statistique intégré lui permettant de convertir le langage naturel en langage de programmation. Cela devient ainsi une routine de reconnaissance vocale. Pour ce faire, le discours est décomposé en minuscules unités. Ensuite, l'ordinateur compare les unités récentes aux unités reçues précédemment. La réponse ou le résultat au format texte détermine, de manière statistique, les mots et phrases qui ont été le plus vraisemblablement prononcés (Frankenfield, 2020).

La deuxième tâche du traitement automatique du langage naturel consiste à reconnaître la classe grammaticale de chaque mot de façon élémentaire, c'est-à-dire s'il s'agit d'un nom, d'un adverbe, d'un verbe ainsi que le temps utilisé etc. Cette étape du processus est appelée étiquetage ou désambiguïsation des catégories de mots. Une fois ces deux tâches réalisées, l'ordinateur est alors en mesure de comprendre le sens du discours émis (Frankenfield, 2020).

Finalement vient la troisième et dernière étape où a lieu la conversion du texte à la parole. À son tour, le langage de programmation informatique est transformé en format audible et transmis à l'utilisateur (Frankenfield, 2020).

Le TALN peut être utile au sein de nombreux secteurs comme la recherche d'informations, l'extraction d'informations ainsi que dans le traitement de la parole. Ce dernier est le plus intéressant dans le contexte étudié. En effet, le TALN intervient dans la reconnaissance vocale dans le cadre de l'utilisation d'un assistant vocal (IONOS Digital Guide, 2019).

2.3.3.3.1 Compréhension du langage naturel

Une fois l'interprétation littérale réalisée par le TALN, vient le tour du CLN d'identifier l'intention sous-jacente de ce qui a été dit. Le CLN est en fait une sous-partie du TALN. Il s'agit de faire comprendre à l'assistant le sens de la phrase et la demande initiée par l'individu (Walch, 2018). Une fois la parole comprise, le CLN va la transformer en donnée traitable.

2.3.3.3.2 La génération en langage naturel

La génération en langage naturel⁴⁵ est également reprise dans le TALN et représente le processus de générer un texte compréhensible sur base de données structurées (Dilmegani, 2021). Cette méthode est davantage utilisée aujourd'hui, en particulier dans les entreprises, puisque ce système permet de transformer automatiquement une quantité importante de données chiffrées en un contenu pouvant être compris universellement (Marr, s.d.).

2.3.3.3.3 L'apprentissage automatique appliqué au TALN

Mis en exergue ci-dessus (voir section 3.3.3), l'apprentissage automatique est une des applications de l'IA qui permet aux machines de répondre de manière autonome sans avoir été préalablement programmées pour réaliser une tâche définie. Les algorithmes d'apprentissage automatique construisent des modèles sur base d'exemples fournis. L'algorithme traduit ses recherches sur échantillons de données fournis en modèles mathématiques. Ce faisant, ils peuvent établir des prédictions pour, à l'avenir, agir sans programmation émanant d'un humain. La machine ayant acquis des connaissances peut ensuite prendre des décisions et agir lorsqu'elle rencontre des demandes/ problèmes divergents (Verani, 2020).

L'apprentissage automatique en question peut être appliqué au traitement naturel du langage naturel.

2.3.3.4 Synthèse vocale (Text-to-Speech)

Cette brique technologique, qui a pour acronyme TTS⁴⁶ en anglais, intervient en dernier lieu dans le processus et permet de lire la réponse textuelle à l'utilisateur (Chapuzet, 2019), en convertissant un texte en un homologue sonore compréhensible (Dosquet et al, 2020, p. 19). Cette réponse, générée grâce à l'IA, se traduit par une voix synthétique en s'adressant à l'humain qui avait généré une requête (Chapuzet, 2019).

⁴⁵ Génération en langage naturel, voir glossaire n°20.

⁴⁶ TTS, voir glossaire n°41.

Ce cycle permet ainsi d’avoir une conversation de bout en bout d’un humain à une machine et inversement. Grâce au STT, la voix synthétique exécute l’action requise, qui débouche sur une réponse de la part du système (Chapuzet, 2019).

2.3.3.5 Phonologie

Afin de comprendre comment une machine est capable d’apprendre à décrypter le langage humain, il est important de connaître les règles de phonologie. Il s’agit de la partie du langage qui traite les sons syllabiques et sous-syllabiques. En effet, chaque langue humaine est en réalité une séquence de phonèmes. Il existe actuellement 14 000 phonèmes possibles mais seulement 40 peuvent être répertoriés dans une langue. Dans la langue française, 36 phonèmes sont identifiés. Il s’agit des sons « ou », « oi », « on », etc. Les phonèmes sont dits consonantiques ou vocaliques selon la mise en jeu de consonnes ou de voyelles. Afin de savoir si un phonème est sourd ou sonore, on analyse l’intervention des cordes vocales lors de la prononciation du mot. Un phonème est dit sourd s’il fait vibrer les cordes vocales. Dans le cas inverse, il est sonore. Les phonèmes sont eux mêmes composés de phones, qui sont des sons élémentaires. On peut en compter jusqu’à 3000 dans une langue, ce qui donne une infinité de combinaisons possibles. On parle également de synapses, qui garantissent que les phonèmes soient séparés lorsque nous parlons (Académie de Caen, s.d. ; Le Cun, 2019, p.232 ; Winston, 2010).

2.3.4 Limites dans le fonctionnement de l’assistant vocal

L’assistant vocal étant encore en plein développement, son fonctionnement doit pouvoir répondre à certaines conditions afin de maximiser son adoption. Celui-ci présente néanmoins un bon nombre de limites qui sont en cours d’être adressées. Parmi les principales, nous identifions les limitations du mode parole, les erreurs de sémantique, le *endpointing*⁴⁸ et la pollution sonore.

2.3.4.1 Limitations du mode parole

Comme mentionné dans la section 2.1.1, le mode parole apporte un caractère naturel, rapide et efficace quant à la communication avec la technologie. Ces aspects doivent cependant être

⁴⁸ Endpointing, voir glossaire n°18.

nuancés par le caractère inhérent de la parole, à savoir la variabilité. Parmi ces limitations, se retrouvent la variabilité de l'environnement acoustique, de l'ambiguïté de la langue ou encore du locuteur (Minker & Néel, 2002).

La première variabilité se réfère aux bruits stationnaires et non stationnaires de l'environnement, relatifs à la pollution sonore (voir 2.3.4.4), alors que la deuxième rejoint les problèmes de sémantique et du *endpointing*, présentés dans les deux sous-sections suivantes respectivement (2.3.4.2 et 2.3.4.3). La variabilité du locuteur, quant à elle, reprend les différences que peuvent avoir une même requête en fonction de l'état émotionnel et physique de l'utilisateur. La complexité de comprendre une même demande suivant la voix d'une personne, son état cognitif ainsi que son débit de parole reste importante (Minker & Néel, 2002).

2.3.4.2 Erreurs de sémantique

Les erreurs de compréhension sémantique constituent une limite dans l'utilisation des assistants vocaux, en particulier dans le processus de CLN. En effet, la sémantique permet de comprendre le sens des mots et du sujet parlés et cette mauvaise compréhension par les assistants vocaux reste encore fréquente.

Nous allons expliquer ce type d'erreur que l'on peut rencontrer lors de l'utilisation d'un assistant vocal à travers un exemple afin d'illustrer de manière concrète le problème de la compréhension sémantique. Demander à son enceinte connectée de changer de musique en disant par exemple « Ok Google, change », au lieu de changer la musique, l'assistant vocal va informer l'utilisateur sur les bureaux de change que ce dernier peut trouver à proximité. Dans cet exemple, le système d'IA est apte à reconnaître la voix mais n'a pas été capable d'interpréter la demande initiale de l'utilisateur. Le problème vient de l'interprétation sémantique de la demande (Ioualalen, 2021).

Les chercheurs sont d'ailleurs toujours attelés à l'amélioration des algorithmes informatiques pour perfectionner leur capacité de compréhension sémantique du langage humain (Ioualalen, 2021).

2.3.4.3 Endpointing

Les assistants vocaux ont encore des difficultés à identifier quand l'utilisateur a fini de parler. Le système de *endpointing* pose problème dans la reconnaissance vocale. En effet, la détection du *endpointing* dans une interface vocale équivaut à l'appui sur le bouton « entrée » sur un clavier, mais dans le cas de l'assistant vocal, l'utilisateur n'annonce pas qu'il a terminé de parler (Black, 2019).

De même, si l'utilisateur marque des pauses lorsqu'il parle à son assistant vocal, ce dernier peut alors considérer ces temps d'interruptions comme un signal signifiant que la requête est finie (Black, 2019). Ce phénomène rend l'utilisation de l'assistance vocale médiocre. Des algorithmes se développent afin de minimiser ce genre de problème et de rendre la communication avec l'assistant vocal de plus en plus humaine (Aleksic et al., 2017).

2.3.4.4 Pollution sonore

Afin d'obtenir une reconnaissance vocale valable, il est important de limiter les erreurs du STT. En fonction de l'environnement dans lequel l'individu se trouve au moment où il parle, il se peut que la pollution sonore soit présente. Le cas échéant, celle-ci peut interférer sur les fréquences sonores enregistrées. Afin d'assurer le bon enregistrement de ces fréquences et, par conséquent, leur fiabilité, plusieurs traitements sont effectués (Chapuzet, 2019). Parmi ces traitements, Aurélien Chapuzet (2019) cite « la normalisation servant à supprimer les pics et les creux dans les fréquences afin d'harmoniser l'ensemble », « la suppression des bruits de fond pour améliorer la qualité audio » ainsi que « la découpe des segments en phonèmes ». Le perfectionnement de ces traitements relatifs à la pollution sonore est indispensable pour garantir un fonctionnement plus optimal de l'assistant vocal.

2.3.5 Le marché des assistants vocaux

Ayant une bonne compréhension de l'assistant vocal ainsi que de son fonctionnement, nous allons maintenant nous intéresser à son marché. L'interface vocale étant l'apanage de l'homme, les assistants vocaux prennent en popularité dans plusieurs domaines. Dans cette partie, nous regarderons aux différents terminaux dans lesquels l'assistant vocal s'immisce, nous

aborderons également les usages principaux de celui-ci au sein des différents domaines et nous nous intéresserons aux secteurs les plus impactés par l'émergence de l'assistant vocal.

2.3.5.1 Les terminaux dotés d'un assistant vocal

De plus en plus de dispositifs connectés sont munis d'un assistant vocal afin de rendre l'interaction entre l'utilisateur et la machine plus fluide et naturelle. L'un des plus courants est l'enceinte connectée, mais il est néanmoins important de préciser que cette dernière est parfois confondue à tort avec l'assistant vocal, alors qu'elle ne constitue que la partie physique et apparente de l'assistant (Dosquet et al., 2020, p.15). En d'autres termes, l'assistant vocal n'est pas une enceinte connectée, mais une enceinte connectée peut disposer d'un assistant vocal (CNIL, 2020). Même si l'enceinte connectée a été l'un des éléments catalyseurs du développement de l'interface vocale, elle n'est pas la seule option possible (Kinsella & Mutchler, 2018). D'après un rapport publié par Hadopi & le Conseil supérieur de l'Audiovisuel⁵⁰ (CSA) (2019), les assistants vocaux étaient, dans un premier temps, uniquement disponibles sur les smartphones et tablettes. Ils ont ensuite été développés davantage afin d'être dorénavant accessibles sur les ordinateurs, télévisions, enceintes connectées, écouteurs, montres ainsi que dans les voitures et sur certains appareils électroménagers.

D'après le graphique proposé par Kinsella & Mutchler (2018) (voir figure 9), établi suite à une enquête réalisée avec un échantillon de 1040 adultes américains, l'enceinte connectée constitue le terminal dominant dans l'utilisation de l'assistant vocal, avec 93% des personnes interrogées disant en posséder ou en avoir déjà utilisé. Le smartphone représente le deuxième terminal sur lequel une grande majorité des répondants utilisent un assistant vocal. Ensuite, la technologie vocale incorporée dans les voitures est utilisée par la moitié des personnes ayant participé à l'enquête. Enfin, l'interaction avec les autres terminaux susmentionnés représentent une plus fine utilisation mais promettent une expansion considérable.

⁵⁰ CSA, voir glossaire n°12.

Proportion de l'utilisation de l'assistant vocal par appareil

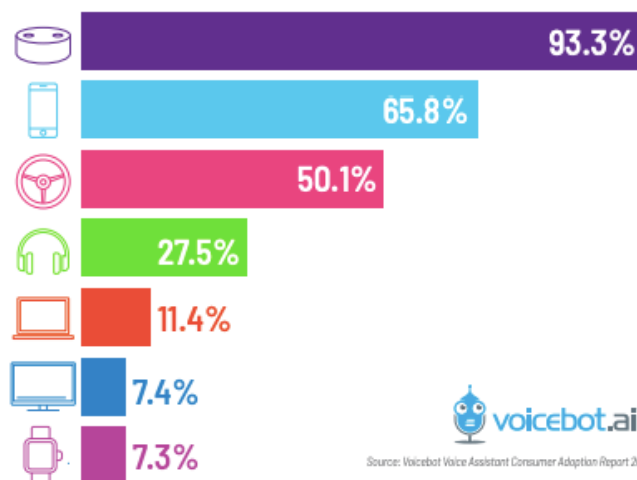


Figure 9: Utilisation de l'assistant vocal sur les différents terminaux
(Source : Voicebot, 2019)

Selon le terminal adopté, les fonctionnalités de l'assistant vocal peuvent différer de manière significative. En effet, concernant l'enceinte connectée, où la voix est en fait le seul mode d'interaction avec ce terminal, les usages sont multiples et souvent reliés à la domotique du domicile. Ils peuvent répondre à une requête de musique, d'informations simples mais également de gérer des objets connectés de la maison comme par exemple des ampoules ou autre (Hadopi & CSA, 2019). Le smartphone est plutôt utilisé pour téléphoner, envoyer des messages, saisir une destination GPS⁵¹, ou encore mettre une alarme (Hadopi & CSA, 2019 ; Kinsella & Mutchler, 2018). Sur les écrans intelligents dont les tablettes et ordinateurs sont munis, les consultations de la météo, du calendrier ou encore des recherches simples sont privilégiées. Par ailleurs, les usages sont plus restreints sur les télévisions, où le démarrage du téléviseur et le changement de chaînes sont pratiquement les seules utilités (Hadopi & CSA, 2019).

Ces utilisations constituent quelques-uns des principaux usages de l'assistant vocal, bien que l'assistant puisse servir à un nombre significatif d'usages ultérieurs. Nous irons plus en profondeur de ces usages dans le point suivant.

⁵¹ GPS, voir glossaire n°22.

2.3.5.2 Les principaux usages

Plusieurs des principaux usages ont été mentionnés dans le point précédent. Par ailleurs, selon Dosquet et al. (2020, p.21), les utilisateurs emploient la technologie vocale principalement pour quatre familles d'usages distinctes. La première englobe les médias et divertissement, dont l'écoute de contenu d'audio est la plus fréquente. La deuxième concerne la recherche d'information. Dans cette catégorie, la demande de météo prône parmi les requêtes demandées ainsi que tout autre renseignement de caractère général. Dans la voiture, ce genre de demandes relève plutôt de la recherche d'adresse, de commerce ou de médecin, etc. La troisième rubrique comprend le contrôle de la maison connectée, c'est-à-dire les usages relatifs à la domotique comme allumer les lumières ou le chauffage par exemple. Enfin, la dernière famille d'usages reprend le commerce vocal afin de réaliser des achats en ligne.

Ces différents usages sont constatés comme restant relativement basiques. Nous remarquons que les assistants vocaux sont sollicités pour faciliter les tâches quotidiennes déjà considérées comme simples plutôt que pour résoudre des requêtes complexes (Verchère, 2018). En effet, ces utilisations satisfont essentiellement à des besoins de nature fonctionnelle, comme le gain de temps ou la commodité d'utilisation, induisant alors un aspect de « gadget » plutôt qu'une réelle nécessité (Hadopi & CSA, 2019). Malgré le fait que les tâches basiques restent la norme, les capacités d'utilisation sont néanmoins grandissantes et certains usages plus sophistiqués commencent à se développer davantage dans plusieurs cadres d'utilisation afin d'offrir une meilleure expérience aux utilisateurs. Ceux-ci ne sont pas encore exploités dans une large mesure, mais ils ne tarderont pas à se généraliser (McCaffrey et al., 2018). Nous verrons ces capacités croissantes dans le prochain point et nous analyserons dans ce mémoire qu'une majorité de ces requêtes ont un avenir prometteur.

De plus, les entreprises dominantes telles qu'Amazon, Apple et Google créent leur propre écosystème⁵² afin de conserver leur indépendance. Le fait d'avoir un écosystème peut être avantageux pour l'utilisateur étant donné qu'il va pouvoir contrôler simultanément depuis différents dispositifs (Bergounhous, 2018).

⁵² Écosystème, voir glossaire n°16.

2.3.5.3 Les cadres d'utilisation

Les assistants vocaux sont majoritairement présents au sein des véhicules, dans les domiciles et, un peu moins répandu, sur les lieux de travail. Ces cadres d'utilisation font l'objet de beaucoup d'avancées étant donné que la voix devient un mode de communication plus pratique que l'interface écrite, rendant alors l'utilisation du vocal inévitable. Les secteurs les plus impactés par l'arrivée de la technologie vocale parmi ces trois cadres d'utilisation ne sont autres que des domaines de la vie de tous les jours, à savoir l'automobile, la santé, l'éducation, les médias et divertissement, les banques et le commerce en ligne (Dosquet et al., 2020, p.29 & 60). Pour notre mémoire, nous nous intéresserons particulièrement aux trois premiers, qui correspondent aux cas d'usages sur lesquels nous avons décidé de nous pencher.

Ces trois secteurs choisis présentent d'importantes opportunités où le vocal pourrait s'avérer révélateur. De fait, les constructeurs automobiles insèrent de plus en plus la technologie vocale dans leurs nouveaux modèles de voitures en la créant en interne, mais établissent également des partenariats avec les géants technologiques comme Amazon (Echo Auto), Apple (Car Play) ou encore Google (Android Auto) afin de permettre aux automobilistes de connecter leur smartphone dans l'habitacle de la voiture. Il sera alors possible d'effectuer toutes sortes de tâches vocalement depuis son véhicule, allant d'utilisations propres à la voiture aux utilisations plus génériques (Dosquet et al., 2020, p.63 ; Hadopi & CSA, 2019).

Concernant le secteur de la santé, l'assistant vocal s'invite dans l'accompagnement quotidien des personnes en situation d'handicap, des personnes âgées et également des personnes dépendantes. L'interface vocale se conjugue très bien avec le spectre du vieillissement de la population et le développement de la maison intelligente. En effet, avoir un assistant vocal dans le domicile peut constituer un réel atout dans l'aide et la sécurité de ces personnes (Dosquet et al., 2020, p.65). La société Libertana⁵³ aux États-Unis a développé une compétence via Alexa d'Amazon visant à permettre aux personnes vulnérables de pouvoir avoir accès à leurs indicateurs de santé et données médicales comme le poids et la pression artérielle, ou encore de bénéficier d'un rappel de prise de médicaments (Regan, 2017). Davantage de tâches pourront être effectuées par l'assistant vocal afin de faciliter la vie et aider les personnes dans le besoin.

⁵³ Libertana, basée à Los Angeles en Californie, est une agence de soins à domicile qui fournit aux familles des soins médicaux et non médicaux.

Même si les avancées sont encore assez limitées, le secteur de l'éducation peut également bénéficier de la technologie vocale. En effet, les assistants vocaux peuvent servir de support dans l'enseignement afin d'aider les étudiants dans leur processus d'apprentissage. Aux États-Unis, Alexa Echo⁵⁴ a été testé dans plusieurs écoles supérieures afin d'aider les professeurs dans certains procédés, ainsi que de permettre aux élèves de demander des informations à l'assistant vocal plutôt qu'à leur ordinateur. De plus, ces dispositifs peuvent être utilisés à des fins d'apprentissage de langues étrangères, car un des enjeux importants des apprenants est de trouver une occasion de parler la langue. À l'heure actuelle, Alexa parle 7 langues, Google Assistant en parle 13 et Siri en parle 21, donnant désormais la possibilité d'interagir avec les utilisateurs désireux d'apprendre une nouvelle langue (Satratzemi & Terzopoulos, 2020).

2.3.6 Sécurité et confidentialité

Bien que les assistants vocaux présentent des caractéristiques attrayantes et utiles, ils posent également plusieurs problématiques et craintes qui peuvent contrer l'adoption de ces technologies. L'un des problèmes principaux à l'utilisation de ces dispositifs à commande vocale est la sécurité des données personnelles. En effet, en raison de son mode de fonctionnement, l'assistant vocal récolte toutes sortes de données de l'utilisateur lorsque celui-ci lui pose des questions ou effectue des achats, représentant un risque que ces données soient exploitées ultérieurement (Hoy, 2018). Au plus le système dispose de données résultant d'un usage fréquent, au plus l'assistant vocal aura une meilleure prédiction de l'intention de l'utilisateur. Cela peut se révéler pratique d'un côté, mais d'un autre côté, cela prouve que les utilisateurs n'ont pas le plein contrôle de leurs données personnelles (Edu et al., 2020). C'est pourquoi les assistants vocaux sont confrontés au Règlement Général sur la Protection des Données (RGPD⁵⁵) (Chapuzet, 2021).

Tout assistant vocal doit respecter le RGPD, qui vise à protéger les données à caractère personnel de tout résident européen. L'une des règles de base sous le RGPD est l'obtention du consentement de l'utilisateur en ce qui concerne la collecte et le traitement des données. L'utilisateur a le droit de savoir quelles données sont détenues par l'assistant vocal, mais

⁵⁴ Alexa Echo est l'enceinte connectée d'Amazon.

⁵⁵ RGPD, voir glossaire n°37.

également les rectifier ou les supprimer (Nicole, 2020). L'idée fondamentale est que la personne qui utilise le système doit pouvoir, au préalable, confirmer ou rejeter l'utilisation de ses données personnelles (Chapuzet, 2021).

En plus de la sécurité, la confidentialité et protection de la vie privée représentent une autre préoccupation majeure pour les utilisateurs d'assistants vocaux. Ces appareils, afin qu'ils soient efficaces et fonctionnels, doivent être à l'écoute en permanence (Hoy, 2018). En effet, ils sont à l'écoute sans arrêt en attendant le mot déclencheur. Il arrive que l'utilisateur prononce involontairement ce mot déclencheur ou tout autre mot semblable, ce qui engendre les assistants vocaux à enregistrer les paroles de l'utilisateur sans que ce dernier le veuille (Edu et al., 2020). Les entreprises qui confectionnent et développent ces assistants vocaux tentent d'être davantage vigilantes et méticuleuses, mais il nécessitera encore d'importantes améliorations afin que ces dispositifs à commande vocale assurent une sécurité et confidentialité maximales (Hoy, 2018).

2.3.7 Le futur des assistants vocaux

La sophistication et la précision des assistants vocaux ont connu une croissance exponentielle dans le courant de ces dernières années. Ces technologies vocales vont connaître un nombre important d'opportunités futures pour lesquelles l'assistant vocal pourra apporter une plus-value dans des domaines primordiaux (Hoy, 2018). En effet, en adoptant une interface qui se rapproche le plus du naturel, mais également en assurant la simplification de la vie de leurs utilisateurs, les assistants vocaux sont amenés à faire partie intégrante des usages du quotidien de tous, de façon graduelle et durable (Alimi, s.d.).

Néanmoins, il n'est pas à négliger que les assistants vocaux nécessitent encore de beaucoup d'améliorations avant d'être adoptés par une grande majorité. Outre les progressions requises dans le mode parole, la compréhension sémantique, la minimisation du phénomène du *endpointing*, ainsi que le perfectionnement de la reconnaissance vocale en général, le principal enjeu sera de parvenir à avoir une vraie conversation avec l'assistant vocal dans un but d'offrir une expérience unique à l'utilisateur (Alimi, s.d.). Les contrôles de sécurité et de confidentialité devront également être renforcés au vu des usages futurs qui nécessiteront d'autant plus de confiance envers les dispositifs à commande vocale (Hoy, 2018).

Si, pour le moment, les possibilités d'utilisation de ces dispositifs à commande vocale sont encore restreintes, la force de l'IA et le développement de ses algorithmes vont toutefois mener vers le déploiement de ces assistants vocaux afin qu'ils parviennent à répondre aux demandes toujours plus complexes dans les prochaines années à venir (IA School, 2020).

CHAPITRE 3: MÉTHODOLOGIE

Après avoir expliqué les deux concepts majeurs, que sont l'intelligence artificielle et les assistants vocaux ainsi que leur lien, nous sommes en mesure de décrire la méthodologie appliquée à notre recherche ainsi que la manière dont nous avons conduit nos entretiens. Nous indiquerons également la structure des chapitres à venir.

3.1 Choix du sujet

Le principal objectif de ce mémoire est l'analyse de trois cas d'usages différents dans lesquels l'assistant vocal apporterait une plus-value. Nous avons choisi de nous concentrer sur les sujets suivants : l'aide à la conduite (1), l'apprentissage d'une langue étrangère (2), et les services de secours à distance (3). Ceux-ci seront décrits en détail dans le chapitre suivant.

Nous avons choisi ces trois cas pratiques d'une part car nous estimons que ceux-ci sont probants à une utilisation accrue de technologie vocale. Cette dernière semble d'ailleurs avoir un potentiel prometteur dans un futur proche. D'autre part, la raison pour laquelle nous avons opté pour des cas pratiques est d'instaurer un cadre à notre recherche. Cela nous a permis de limiter nos recherches et nous concentrer sur ces secteurs identifiés. De plus, nous désirions analyser des cas d'usages bien distincts et indépendants les uns des autres afin d'évaluer si ces technologies vocales ont réellement des possibilités de succès dans différents domaines.

Le choix de ces trois cas d'usages précis s'est fait suite à une lecture rigoureuse d'articles scientifiques ainsi qu'un brainstorming avec notre promoteur. En effet, il en est ressorti que l'assistance vocale occupe une place grandissante dans les voitures, lieux de travail et domiciles. De là, nous avons réfléchi aux situations dans lesquelles les assistants vocaux acquièrent des opportunités de développement dans les trois différents domaines. Ainsi, nous avons retenu les trois cas d'usage énumérés précédemment.

3.2 Choix de l'étude

Il existe deux types de méthodologie, à savoir la méthodologie qualitative et la méthodologie quantitative. L'étude qualitative repose sur une collecte de données non quantifiables, réalisée

sur base d'entretiens individuels ou collectifs avec des experts ou simplement par de l'observation. Elle a pour but de comprendre et d'évaluer des comportements pour ensuite analyser et interpréter les données récoltées (Claude, 2019 ; Jacquemin, 2017). La méthodologie quantitative, quant à elle, vise à prouver un phénomène plutôt que de l'interpréter, ainsi que d'en tirer des résultats qui sont statistiquement mesurables (Claude, 2019).

L'option que nous avons retenue pour élaborer la partie pratique de notre mémoire est celle de la recherche qualitative. Il nous paraissait plus logique dans notre cas de consulter l'avis d'experts plutôt que de rassembler des données chiffrées. Cette recherche s'est menée sur base d'interviews d'experts présents dans les trois secteurs de notre recherche. Nous avons opté pour la méthode d'entretiens semi-directifs, que nous détaillerons dans la section suivante. En effet, pour chaque cas de figure, nous avons élaboré un guide d'entretien et défini une hypothèse de recherche. Ces interviews ont été réalisées en visioconférence et ont duré entre 30 minutes et 1h30.

3.3 Description du design de la recherche

Dans le cadre d'une étude qualitative, nous passons par trois étapes, à savoir le choix du terrain et la constitution de l'échantillon (1), la collecte de données (2) et finalement l'analyse de données (3).

Afin de conduire nos interviews, la constitution de notre échantillon a été une tâche primordiale. Il a été nécessaire d'avoir un échantillon de personnes assez diversifié pour pouvoir rassembler des avis divergents. La taille de notre échantillon n'a pas été défini de prime abord. Cela s'est fait naturellement en fonction des profils trouvés et recommandés par certains intervenants. Notre échantillon se compose de 10 personnes. Chacune d'elles présente un lien avec les assistants vocaux, que ça soit par l'expertise ou par leur utilisation.

La technique de collecte de données que nous avons suivie est la technique des entretiens individuels semi-structurés. Cette forme d'entretien est particulièrement adaptée pour connaître les positions par rapport à un sujet. Cela nous a permis de garder une certaine flexibilité lors des entretiens et d'aborder certaines thématiques en fonction des propos de la personne. Cette technique a pour dessein de coproduire les données récoltées et d'apporter une certaine richesse

d'information. Ainsi, cette technique s'est avérée être la plus propice au regard de notre recherche (Decrop & Masset, 2019).

Pour ce faire, nous avons créé un guide d'entretien semi-directif⁵⁶ pour chaque cas. Nous avons suivi la structure classique recommandée par Pleyers (2020), à savoir : introduction, warm-up, focus, approfondissement et finalement, conclusion. À l'instar d'un entonnoir, cette structure permet de traiter des éléments généraux pour finir par des questions plus précises. Nous avons également inclus au guide le message d'introduction adressé à chaque intervenant.

Nous avons clôturé nos interviews une fois que nous avons estimé avoir atteint un critère de saturation acceptable. Au fur et à mesure des entretiens, nous avons enrichi nos données. Les premiers entretiens nous ont également permis d'améliorer notre guide d'entretien pour les interviews qui suivaient. Les guides ont également pu être façonnés pour s'adapter au mieux au profil interrogé.

Bien que nous sommes parvenues à interroger des profils différents, et ce pour chaque cas d'usage, nous avons également essuyé de nombreux refus. Tout cas d'usage confondus, nous avons contacté 45 personnes au total⁵⁷. Cela représente un taux de succès de nos demandes d'interview de 22,22%. Toutes les données collectées au cours des 10 entretiens se sont avérées être exploitables. Nous tenons d'ailleurs à remercier chacun des intervenants d'avoir participé à l'interview et montré leur intérêt quant au sujet de notre recherche. Ces derniers sont présentés ainsi que leur fonction respective dans un tableau⁵⁸.

3.4 Analyse de données

Nous avons clôturé les interviews d'experts le 13 juillet 2021. Celles-ci constituent la matière première de notre recherche. Chaque interview a fait l'objet d'enregistrement et a été ensuite transcrit⁵⁹, comme recommandé par le Professeur Christophe Lejeune. Celui-ci prône une retranscription mot à mot afin d'éviter une sélection des idées selon les à priori des chercheurs.

⁵⁶ Guides d'entretien, voir annexe n°4

⁵⁷ Tableau des personnes contactées en vue de les interviewer, voir annexe n°3

⁵⁸ Tableau de profils des experts en annexe n°6

⁵⁹ Voir transcription des interviews en annexe n°6

Nous avons ensuite dressé des tableaux de synthèse⁶⁰, par cas d'usage, afin de mettre en perspective les thématiques les plus pertinentes discutées au cours des entretiens. Ces tableaux à double entrée reprennent, verticalement, les principaux thèmes discutés lors des interviews. Horizontalement, on retrouve une case pour chaque intervenant. Ces matrices ne sont pas exhaustives quant aux informations récoltées. Cependant, elles nous ont permis de les synthétiser et de réaliser un plan pour la suite de notre analyse.

Sur base de ces tableaux, nous avons analysé les données récoltées et nous sommes parvenues à des conclusions confirmant ou réfutant l'hypothèse de chaque cas d'usage. L'analyse sera structurée de manière à prendre en compte chaque facteur entrant en considération avec l'adoption de l'assistant vocal.

Pour chacun des cas d'usage, nous avons étudié l'offre et la demande de la technologie vocale dans les domaines respectifs. En effet, notre analyse se structure comme suit : d'une part nous avons analysé les facteurs clés de succès que l'assistant vocal peut offrir au sein des cas d'usage. D'autre part, nous avons examiné les éléments que les utilisateurs demandent et jugent nécessaires afin d'adopter l'usage de la technologie vocale, mais également les facteurs pouvant empêcher cette adoption de se produire. Sont aussi abordées les prévisions futures que les experts imaginent pour les assistants vocaux dans les trois cas d'usage. Une conclusion est élaborée à la fin de chaque cas, résumant les éléments clés retrouvés dans l'analyse et identifiant la confirmation ou le rejet de l'hypothèse de recherche respective à chaque cas d'usage.

⁶⁰ Voir tableau d'analyse des données collectées en annexe n°5

CHAPITRE 4: PARTIE EMPIRIQUE

Dans ce chapitre, nous allons analyser en détail chacun de trois cas d'usages choisis et parvenir à des hypothèses de recherche qui seront alors confirmées ou réfutées dans le cadre de la partie analyse de ce mémoire. Cette partie analyse recense les informations recueillies lors des entretiens menés avec les experts.

4.1 Cas d'usage 1 : L'aide à la conduite

4.1.1 Description du cas

L'industrie automobile a progressé ces dernières années en matière de commodités et de connectivité, notamment grâce aux assistants vocaux. L'assistant vocal, tel que décrit dans notre revue de littérature, s'invite dans l'habitacle des véhicules. Le nombre de personnes utilisant un assistant vocal ne cesse de croître, tout comme les constructeurs automobiles qui intègrent davantage cette nouvelle technologie dans le but d'offrir une expérience client toujours plus personnalisée. Ces assistants vocaux peuvent soit originellement se trouver dans le véhicule, c'est-à-dire en embarqué⁶¹, soit depuis les smartphones des consommateurs, et permettent tant au conducteur qu'aux passagers de bénéficier d'une multitude de services, sans pour autant retirer leurs mains du volant et leurs yeux de la route (SoundHound, 2020).

En effet, davantage de voitures sont dotées d'un assistant vocal, qui permet au conducteur et aux passagers d'accéder à des fonctions de divertissements et d'informations. Selon Statista (2021), les fonctions les plus populaires auprès des consommateurs sont celles liées à la conduite et au rôle du passager, à savoir notamment la lecture de musique ou encore la demande d'itinéraire. Les actions de composer un numéro pour appeler ou envoyer des messages sont également des tâches clés que le conducteur aborde progressivement avec la voix et non plus avec le toucher (Kinsella, 2020). Plus élaboré, il y a aussi l'interaction avec les appareils domestiques du domicile, c'est-à-dire le système de domotique, qui prend de l'ampleur dans l'utilisation de l'assistant vocal dans la voiture. En effet, accéder au thermostat, aux lumières

⁶¹ Système embarqué, voir glossaire n°39.

ou encore aux systèmes de sécurité du domicile peut apporter une valeur ajoutée à l'expérience du conducteur (SoundHound, 2020).

D'autres cas d'usage plus complexes d'une assistance vocale au sein des voitures peuvent comprendre des services plus contextuels et personnalisés comme le fait de recommander au conducteur de s'arrêter à un certain endroit pour faire ses courses, en fonction de l'heure, du lieu, du trafic et des habitudes du consommateur connues par l'assistant vocal (Helmer, 2021). D'autres actions davantage implémentées dans les voitures sont les actes de recommander une station-service proche et peu chère lorsque l'assistant détecte que le niveau de carburant est faible, ou encore de proposer au conducteur de prendre une pause quand la voiture constate une longue durée de conduite (Minker et al., 2019). Nous observons qu'il existe dès lors un grand nombre de possibilités d'usage d'un assistant vocal au sein des voitures.

Bien qu'il ait parfois des allures d'accessoire, au vu des principaux usages décrits dans les paragraphes précédents, en voiture, l'assistant vocal présente des avantages pour le conducteur ainsi que pour les passagers. Ceux-ci utilisent davantage la technologie vocale pour effectuer des tâches pour lesquelles ils doivent généralement regarder et utiliser leur téléphone et/ou tableau de bord du véhicule. L'assistant vocal est alors exploité en partie afin de renforcer la sécurité du conducteur et de le distraire le moins possible de la route (Kinsella, 2020).

Alors que ces nouveaux usages contribuent à nous rendre bon nombre de tâches plus faciles, il reste cependant des risques liés à nos usages, tant professionnel que personnel, dont il est important de prendre conscience et qui peuvent constituer des freins à l'adoption de cette intelligence artificielle dans les véhicules par les consommateurs.

4.1.2 Hypothèse de recherche

Sur base de nos connaissances et intuition au sujet de l'assistance vocale dans les voitures, nous pensons que, malgré les risques qu'il encourt, l'assistant vocal promet une réelle plus-value pour l'expérience des automobilistes. Il y a en effet, selon nous, un véritable potentiel et, d'ici peu, l'assistant vocal qui s'invite dans l'habitacle des véhicules devrait être omniprésent. Ainsi, nous sommes arrivées à l'hypothèse suivante pour ce cas d'usage :

L'assistant vocal apporte une plus-value en termes de sécurité et d'expérience aux conducteurs de voitures qui en sont équipées.

4.1.3 Analyse scientifique sur base des interviews

Afin de pouvoir confirmer ou réfuter l'hypothèse susmentionnée, nous allons analyser les avis de chacun des experts que nous avons interviewé pour ce cas d'usage. Les quatre personnes que nous avons eu la chance de rencontrer ont amené des opinions et perspectives bien distinctes les unes des autres. Il convient de prendre en compte les éléments qu'offre l'assistant vocal dans le secteur automobile mais également les aspects attendus par les conducteurs, ainsi que ceux qui les retiennent d'adopter cette interface vocale.

4.1.3.1 Facteurs clés de succès (Offre)

Dans cette section, nous allons nous intéresser à l'offre de la technologie vocale en voiture, au sens où nous allons couvrir l'analyse des éléments que l'assistant vocal apporte en conduisant, selon les propos des différents intervenants pour ce cas d'usage premier.

Selon Oxana Gouliaéva, la présence d'un assistant vocal dans l'habitacle du véhicule répond à deux contraintes principales, à savoir le temps pour réaliser une tâche ainsi que la contrainte mains libres. L'utilisation du vocal représenterait un gain de temps important et non négligeable aux yeux des utilisateurs. De plus, la contrainte des mains est résolue grâce au vocal, permettant au conducteur de pouvoir effectuer des tâches tout en gardant ses mains sur le volant. Soledad Bedu rejoint cette idée en suggérant que l'assistant vocal assure une certaine rapidité et facilité. Pour illustrer cette idée, celle-ci évoque l'exemple d'un message WhatsApp, en expliquant que, grâce à Android Auto⁶², l'assistant embarqué lui propose de lui lire le message et lui suggère ensuite d'y répondre par la voix. Cela résulte en un acte rapide, pratique, facile, tout en ayant gardé ses mains sur son volant.

⁶² Android Auto est une application mobile développée par Google pour reproduire les fonctionnalités d'un appareil Android, tel qu'un smartphone, sur l'unité principale d'information et de divertissement du tableau de bord d'une voiture (Wikipedia, 2021).

La même intervenante soutient l'idée que l'assistant vocal aide le conducteur à avoir une conduite plus sécuritaire, plus confortable, et également plus personnalisée. Un autre expert, Jean-François Bonastre, affirme que ces dispositifs vocaux peuvent représenter une aide pour les petits usages tels que téléphoner, mais doute que ce soit déjà le cas pour les usages plus importants. Celui-ci avance qu'il y a de gros problèmes d'ergonomie et de performance, qui doivent être améliorés afin de réellement représenter une plus-value pour le conducteur. Madame Bedu a également élaboré cette idée en mentionnant que l'assistant vocal apportera une plus-value s'il est initialement conçu correctement et s'il a été réfléchi afin que son utilisation soit intuitive pour l'utilisateur.

Outre ces plus-values élaborées ci-dessus, la majorité des quatre intervenants pensent que l'assistant vocal peut, d'une certaine manière, apporter une sécurité supplémentaire à la conduite mais restent tout de même sceptiques sur certains aspects. En effet, Ariane Nabeth-Halber et Soledad Bedu estiment, toutes deux, que tout ce qui peut permettre de faire des requêtes à la voix sans détourner le regard de la route ni enlever ses mains du volant est intéressant et peut théoriquement apporter une certaine sécurité. Ensuite, Oxana Gouliaéva considère que tout dépendra des usages ainsi que de la performance de l'assistant vocal, mais que celui-ci n'apporte pas forcément une sécurité additionnelle. Quant à Jean-François Bonastre, ce dernier approuve le fait que l'assistance vocale peut permettre d'éviter les actions dangereuses telles que composer un numéro ou rédiger un message sur un écran, mais n'est pas convaincu de la réelle sécurité que cette technologie peut apporter dans la conduite en elle-même. Selon lui, la communication par la voix est intéressante pour sélectionner et recevoir de l'information, mais il ne faut pas s'attendre à pouvoir contrôler sa voiture vocalement.

Soledad Bedu ajoute cependant que lorsque les assistants vocaux pourront agir de façon proactive, par exemple, en proposant de chercher une station d'essence proche lorsqu'ils ont détecté un faible niveau de carburant, le côté sécuritaire sera renforcé étant donné que la technologie ira au-delà des besoins de l'utilisateur. Par ailleurs, la détection des états du conducteur, tels que la fatigue, pourrait avoir un impact positif en termes de sécurité sur sa conduite puisque l'assistant vocal réagira en diffusant de la musique pour le réveiller ou simplement en lui conseillant de s'arrêter.

Dès lors, pour chacun des quatre intervenants interviewés pour ce cas d'usage, tous sont d'avis que l'assistant vocal est en mesure d'apporter, d'une manière ou d'une autre, une valeur ajoutée dans la voiture. Ils restent toutefois peu persuadés de sa capacité à apporter une réelle sécurité pour la conduite.

4.1.3.2 Risques

La présence d'assistant vocal dans les voitures a du positif mais présente également un certain nombre de risques qui ne sont pas à sous-estimer. Nos quatre répondants sont unanimes pour affirmer que l'assistant vocal occupe une charge cognitive conséquente au volant. L'interaction vocale simplifie les tâches mais le fait de discuter dans la voiture distrait tout de même le conducteur. Ariane Nabeth-Halber cite que le vocal peut en effet détourner l'attention en projetant l'utilisateur dans un lieu parallèle, réduisant alors sa concentration sur la route. Soledad Bedu ajoute que si l'assistant vocal ne comprend pas la requête demandée, et que les actions ne se font alors pas de façon intuitive, le conducteur sera contraint à réfléchir davantage et cela peut constituer une charge plutôt qu'une aide. La charge mentale de la communication risque d'absorber une partie trop importante de la capacité cognitive, qui devrait être entièrement consacrée à la conduite, selon Jean-François Bonastre.

D'autres risques relatifs à l'assistance vocale dans l'habitacle des voitures sont mentionnés par certains intervenants. Ariane Nabeth-Halber souligne que le système de *endpointing* demande un certain effort quand il est mal conçu, car le conducteur doit se concentrer pour faire sa requête dans un temps imparti afin d'éviter que l'assistant vocal ne l'interrompe. Cette concentration peut alors distraire le conducteur, rejoignant en quelque sorte le premier risque mentionné ci-dessus. Jean-François Bonastre, quant à lui, relève qu'un des plus gros dangers des dispositifs vocaux dans la voiture est qu'il dispose de la localisation du conducteur en permanence et que ce dernier est tracé partout en termes de GPS. Enfin, Soledad Bedu complète l'idée qu'elle a émise au sujet des assistants vocaux proactifs (voir 4.1.3.1) en ajoutant qu'aller au-delà des besoins de l'utilisateur peut apporter une sécurité, mais qu'il est cependant important de ne pas aller au-devant de tous ses besoins, au risque de devenir trop envahissant et intrusif, et ainsi de causer une distraction supplémentaire pour le conducteur.

Le risque de perte de concentration en raison d'une charge cognitive trop importante apportée par l'assistant vocal est un danger qui, selon nos quatre interviewés, est à prendre en considération.

4.1.3.3 Adoption de l'assistant vocal (Demande)

Il est primordial d'examiner ce qui encourage et ce qui retient les conducteurs à adopter l'assistant vocal pendant leur conduite. Nous allons analyser ce que nos interviewés suggèrent pour ces aspects, et il convient également de regarder quel genre de technologie vocale est le mieux adapté à s'inviter dans les véhicules, par qui il doit être développé ainsi qu'aux fonctionnalités les plus appropriées.

4.1.3.3.1 Facteurs clés d'adoption

Lorsqu'une nouvelle technologie apparaît sur le marché, elle réclame toujours un temps d'adaptation. Il y a des conditions que l'assistant vocal doit remplir afin de répondre à la demande des utilisateurs. Notre première interviewée, Oxana Gouliaéva, est d'avis que l'assistant vocal doit être performant tant dans sa compréhension envers ce que l'utilisateur lui demande, que dans la pertinence de la réponse fournie ou de l'action effectuée par l'assistant. En effet, si l'assistant vocal présente des difficultés à comprendre les requêtes ou à y répondre et à agir en conséquence, l'utilisateur n'y verra plus d'intérêt à l'adopter. Ariane Nabeth-Halber estime que l'adoption de cette technologie en voiture se réalise sur base des *business models*⁶⁴ des développeurs d'assistants vocaux. En effet, cette experte a constaté que les entreprises ayant un *business model* initialement compatible avec l'expérience que les utilisateurs cherchent à avoir en voiture, seront plus directement adoptées. Il ne s'agit donc pas seulement de disposer des bonnes technologies, mais également d'avoir un *business model* adéquat. Madame Gouliaéva justifie cette même idée en affirmant que les utilisateurs sont plus stimulés par l'expérience offerte dans la voiture que par la qualité de la conduite.

Soledad Bedu évoque plusieurs idées concernant les facteurs qui peuvent favoriser l'adoption de l'assistance vocale dans la voiture. D'une part, l'assistant vocal doit fournir une utilisation plus simple qu'au toucher. Le cas échéant, l'utilisateur n'adoptera pas l'option vocale et restera

⁶⁴ Business model, voir glossaire n°8.

sur l'utilisation standard. D'autre part, cette experte estime que l'utilisation doit être intuitive et devrait pouvoir assurer un dialogue continu entre l'utilisateur et la machine, ne nécessitant pas de regarder l'écran, ce qui reste encore à l'heure actuelle une lacune. Dans la même direction, madame Bedu défend l'idée que les conducteurs n'adopteront les assistants vocaux dans les voitures que s'ils sont pratiques et faciles d'utilisation. Elle soutient que les applications comme Android Auto et Car Play⁶⁵ répondent à ce genre de critères et dès lors attirent les conducteurs à les adopter pendant leur conduite.

Nos intervenants ont décrit différents éléments qu'ils estiment fondamentaux pour assurer une adoption de la part des conducteurs. En revanche, les facteurs susceptibles de freiner les utilisateurs à adopter cette technologie ne sont pas minimes.

4.1.3.3.2 Facteurs limitant l'adoption

Bon nombre de facteurs viennent encore à l'encontre de l'adoption de dispositifs vocaux dans la voiture. Les personnes interrogées ont des visions diverses sur le sujet.

Premièrement, Oxana Gouliaéva vient balancer ce qu'elle mentionnait envers les facteurs à l'adoption en termes d'expérience au sein de la voiture et ainsi estime que les critères de choix peuvent parfois représenter des limites à l'adoption de dispositifs vocaux. En effet, certains vont peut-être préférer le design à l'expérience et vice versa, et choisiront la voiture ainsi que l'assistant vocal en fonction de ces critères de sélection. Si l'assistant vocal ne rencontre pas ces éléments, l'adoption sera faible pour les personnes ayant ces mêmes besoins. Cette intervenante ajoute également que si le *digital nativity*⁶⁶ (souvent lié à l'âge) et la région du monde dans laquelle certaines personnes se trouvent sont défavorables aux avancées technologiques, l'adoption de l'assistant vocal sera moindre. De fait, si l'attitude générale de cette population est peu propice aux changements technologiques, et que la présence d'acteurs vocaux est faible, cette communauté ne sera probablement pas encline à recourir à ce genre de systèmes innovants. Ensuite, Ariane Nabeth-Halber soutient son opinion initiale concernant le *business model* des technologies vocales en insistant que si le *business model* n'est pas en

⁶⁵ CarPlay est une solution d'accès aux téléphones mobiles/smartphones à l'intérieur d'un véhicule qui permet aux individus d'utiliser leur iPhone dans leur voiture en utilisant le système de contrôle et d'affichage du véhicule. Il s'agit d'un produit propriétaire d'Apple qui offre une fonction d'activation vocale pour accéder à l'iPhone dans un véhicule (Techopedia, 2014).

⁶⁶ Digital nativity, en français : « Nativité numérique », voir glossaire n°14.

adéquation avec la demande et le besoin de l'utilisateur, l'assistant vocal a peu de chances d'être adopté. Il est évident que l'offre doit rencontrer la demande pour avoir une place sur le marché.

Concernant Jean-François Bonastre, ce dernier affirme qu'il y a des réticences techniques de la part des utilisateurs, car il y a la possibilité que certains éléments ne fonctionnent pas ou n'apportent rien de plus qu'au toucher. En effet, il y a un manque d'intérêt envers l'assistant vocal lorsque la question se pose de savoir s'il y a réellement une praticité grâce à la voix. Cet expert donne l'exemple de l'ouverture et la fermeture des fenêtres. Les boutons étant à portée du conducteur, demander à la voiture d'ouvrir la fenêtre n'est pas forcément plus simple. La nécessité et l'utilité de l'assistant vocal doivent alors être évidentes aux yeux des conducteurs afin d'assurer son adoption. Monsieur Bonastre estime dès lors que le vocal tente de s'immiscer dans un trop grand nombre de domaines et d'usages. Il serait certainement plus largement adopté s'il concernait un nombre plus restreint et plus ciblé d'applications. Ce manque d'usages renforcés représente, pour lui, un des freins à l'adoption.

L'un des autres points relevés par monsieur Bonastre concernant les facteurs limitant l'adoption du vocal est celui de la préservation des données personnelles. Il estime qu'il y a une certaine réticence mais que celle-ci n'est pas encore omniprésente. Soledad Bedu, quant à elle, considère que le fait d'ignorer la façon dont les données personnelles sont utilisées représente une crainte majeure auprès des utilisateurs. Lorsqu'il a été découvert qu'Amazon et Google enregistraient les données afin d'améliorer la performance de leurs systèmes, les utilisateurs se sont conscientisés et se sont montrés réfractaires à l'utilisation des assistants vocaux de ces marques. La transparence relative à la gestion des données personnelles de la part des créateurs d'assistants vocaux est essentielle, sans quoi beaucoup de conducteurs n'adopteront pas la technologie vocale par crainte de ne pas avoir connaissance de la façon dont leurs données sont utilisées. Cependant, rien n'empêchera le fait que ces assistants vocaux soient à l'écoute en permanence puisqu'ils attendent de détecter le *wake word*. Il convient alors pour les utilisateurs d'avoir conscience des limites de ces technologies, mais les personnes restent toutefois méfiantes, notamment lorsqu'il s'agit d'avoir des conversations privées en présence d'un assistant vocal.

De plus, madame Bedu est convaincue que les adultes appréhendent l'utilisation d'un assistant vocal et de ce fait évitent d'en adopter. En effet, parler à une machine ne constitue pas une

communication naturelle, entraînant la peur de faire des erreurs et d'échouer lors de la communication avec l'assistant, en particulier en présence de personnes tierces. Cette experte nous explique que lorsque l'assistant vocal ne comprend pas la requête, que ce soit en voiture ou autre, les utilisateurs estiment que c'est de leur faute et ne comprennent pas ce qu'ils ont pu dire ou faire de mal. Leur réaction est alors de surarticuler ou de parler plus fort, ne solutionnant pas l'incompréhension de la machine étant donné que le problème provient de la reconnaissance vocale du système. Cette prise en main de l'outil peut parfois paraître compliquée aux yeux de certaines personnes et l'échec de requêtes peuvent rebuter l'utilisation de ces appareils vocaux.

En lien avec cette idée précédente, Soledad Bedu évoque également le problème des accents lors de l'utilisation d'assistants vocaux dans une langue étrangère par l'utilisateur. Le risque de ne pas être compris est plus important étant donné que le système vocal ne couvre pas tous les différents accents possibles d'une même langue. Afin de permettre une compréhension maximale, un nombre important de données est requis. Cependant, les systèmes n'en disposent pas encore à l'heure actuelle. Par conséquent, s'adresser à un assistant vocal qui ne comprend pas les différents accents ou même certaines requêtes, peut entraver l'adoption de la part des utilisateurs, que ce soit dans la voiture ou même dans d'autres contextes.

Les quatre interviewés de ce cas d'usage avaient des avis différés et variés sur la question des facteurs limitant l'adoption d'assistants vocaux lors de la conduite. Tous pensent qu'il y a encore beaucoup d'aspects venant en travers de l'adoption de cette technologie mais restent convaincus que si ces freins sont traités et améliorés, l'adoption se développera.

4.1.3.3.3 Adoption de technologie embarquée ou externe

Les assistants vocaux au sein des voitures peuvent faire objet d'une technologie embarquée initialement dans le système du véhicule ou depuis une source externe comme les smartphones des conducteurs. La complémentarité des deux méthodes est également une alternative de plus en plus répandue. Chacune des quatre personnes interrogées détient sa propre opinion sur le sujet, mais il semblerait que l'assistant vocal externe ou la complémentarité entre les deux formules prennent le dessus sur la technologie embarquée, selon la plupart des propos.

Oxana Gouliaéva mentionne que les constructeurs automobiles cherchent à intégrer l'assistance vocale mais ne donne pas plus de détails en ce qui concerne son opinion sur le fait d'avoir un assistant vocal embarqué ou externe. Jean-François Bonastre est d'avis que l'intégration de l'assistant vocal au sein des véhicules est relativement peu performante. Il considère qu'il y a un conflit entre le système d'information des voitures et l'assistant vocal car les assistants intégrés n'évoluent pas et ne se mettent pas à jour. Son opinion serait de recourir aux assistants vocaux externes qui soient identiques et compatibles à chaque voiture. En revanche, en prenant en considération l'aspect acoustique et l'aspect application du vocal dans la voiture, la complémentarité de l'assistant embarqué et de l'externe serait plus appropriée. En effet, pour la partie acoustique, la neutralisation du bruit doit être établie de manière subtile. Pour cela, les applications et développements spécifiques des constructeurs automobiles sont intéressants. Concernant la partie application, par exemple si le conducteur souhaite réserver un restaurant, monsieur Bonastre juge que l'utilisateur a plutôt intérêt à utiliser les applications standards provenant d'assistants externes. Nous constatons que son avis est départagé en fonction des facteurs qui interviennent.

Soledad Bedu est également du même point de vue que monsieur Bonastre, mais apporte des éléments d'explications supplémentaires. Cette experte estime que la façon la plus complète d'avoir une bonne communication vocale dans sa voiture est la complémentarité entre la technologie intégrée dans le véhicule et la technologie dans le Cloud⁶⁷, c'est-à-dire la technologie externe venant entre autres de géants technologiques comme Google ou Amazon, ainsi que son système de *back-up*⁶⁸. Certaines fonctionnalités ne sont pas implémentées en embarqué mais le sont dans le Cloud, et vice versa. L'utilisateur doit alors comprendre que les résultats d'une même requête peuvent différer suivant le système utilisé, selon que ce soit l'assistant vocal intégré ou externe qui répond. En effet, s'il y a une perte de 4G⁶⁹, la technologie embarquée ne répondra plus, d'où l'utilité du système de *back-up* avec le Cloud. Pour cette raison, madame Bedu considère que disposer des deux options est plus judicieux afin d'avoir une réserve. Par ailleurs, elle pense que la technologie vocale externe à la voiture permet une meilleure reconnaissance vocale que la technologie embarquée. Dans sa vision, il est alors nécessaire de disposer des deux car chaque option apporte des caractéristiques complémentaires.

⁶⁷ Cloud, voir glossaire n°10.

⁶⁸ Backup ou « sauvegarde » en français, voir glossaire n°6.

⁶⁹ 4G, voir glossaire n°1.

Toujours dans cette idée de complémentarité, Ariane Nabeth-Halber prend l'exemple de Faurecia, fournisseur automobile, qui expérimente également le principe des deux options. L'entreprise française a développé un assistant vocal spécialisé mais fait également recours à un assistant vocal externe pour d'autres fonctions. Une complémentarité entre deux assistants vocaux est alors la meilleure option selon cette experte.

La majorité des personnes interrogées est plutôt persuadée que la complémentarité entre la technologie embarquée et la technologie externe est le moyen le plus complet et le plus efficace d'établir une communication vocale dans la voiture, tout en soulignant que la technologie externe est susceptible d'être plus performante sur le plan de la reconnaissance vocale.

4.1.3.3.4 Adoption de technologie développée par les constructeurs ou acteurs technologiques

Après avoir recueilli l'avis des personnes interrogées sur le choix d'une technologie embarquée ou externe, il est tout aussi indispensable de savoir s'il est plus avantageux pour les constructeurs automobiles de développer leur propre assistant vocal ou d'utiliser la technologie d'autres acteurs. Pour ce sujet, la variété des opinions demeure.

Ariane Nabeth-Halber, dans son exemple évoqué avec Faurecia, estime que l'assistant développé internement pourrait se connecter avec un assistant vocal externe tel qu'Alexa par exemple. La complémentarité des deux, tout comme le point précédent, reste l'option la plus adéquate pour cette intervenante. Oxana Gouliaéva est également légèrement partagée mais considère toutefois que les constructeurs de voitures ont plutôt intérêt à développer leur propre assistant. Elle estime que rien ne les empêche d'expérimenter un partenariat avec les géants technologiques car ceux-ci sont plus performants étant donné qu'ils ont le Cloud, leur permettant d'affiner les données et de perfectionner les algorithmes pour assurer un fonctionnement optimal de l'assistant vocal. Cependant, afin de conserver une relation directe avec le conducteur, le développement d'un assistant vocal propre à la marque de voiture peut s'avérer plus intéressant.

Soledad Bedu partage l'opinion de l'experte précédente, à savoir de développer la technologie en interne, au sein de l'entreprise automobile. Suite aux craintes de l'utilisation des données

personnelles par Google et Amazon (mentionnées dans le point 4.1.3.3.2), madame Bedu évoque que le système natif⁷⁰ peut alors être envisagé. Néanmoins, elle soulève trois facteurs dont pourrait dépendre ce choix. Le premier réside dans les moyens financiers de chaque constructeur. S'il s'agit de constructeurs dont les moyens sont limités, investir dans le développement d'un nouvel assistant vocal peut paraître compromis. Dans ce cas, recourir aux assistants vocaux des grands acteurs semble être la meilleure solution. Pour les constructeurs automobiles qui ont les moyens d'investir, créer son propre système vocal qui fait réellement partie de la voiture est un avantage. Ensuite, le deuxième élément entrant en compte est la clientèle. Certaines marques de voitures ont une clientèle plus âgée n'étant pas apte à utiliser les assistants vocaux externes aux véhicules. Il peut aussi y avoir des marques plus familiales dont les clients disposent d'un Google Home ou d'un Amazon Echo⁷¹ dans leur domicile. Dans ce contexte, il est préférable pour les constructeurs d'utiliser les assistants vocaux développés par ces acteurs afin de faciliter l'usage pour les conducteurs. Le dernier aspect consiste dans l'image de marque. Une marque automobile disposera probablement d'une différente vision perçue de la part des utilisateurs qu'une autre, pour laquelle les constructeurs doivent en tenir vigueur.

Concernant Jean-François Bonastre, ce dernier émet une opinion différente et pense qu'il pourrait y avoir une entente entre plusieurs constructeurs. En effet, il est d'avis que les assistants vocaux peuvent être définis de manière transversale aux marques, c'est-à-dire qu'ils ne doivent pas forcément être développés par chaque marque automobile ni par une multinationale telle que Google. D'une part, il y a une certaine réticence à utiliser Google en ce qui concerne les données personnelles, et d'autre part il suggère qu'un assistant vocal ne doit pas être dédié à une marque particulière de voiture. Ce qui serait intéressant selon lui, c'est effectivement d'avoir plusieurs acteurs spécialisés dans la voiture qui conçoivent un assistant vocal pouvant être compatible dans différentes marques de voitures.

Chacune des quatre personnes que nous avons interviewées émet un point de vue différent des autres, de sorte que chaque perspective présente un intérêt tout aussi important.

⁷⁰ Système natif ou système embarqué

⁷¹ Enceintes connectées de Google et Amazon respectivement.

4.1.3.3.5 Adoption de fonctionnalités spécifiques ou génériques

Outre la disposition d'un assistant intégré ou externe au sein du véhicule, développé par les constructeurs ou autres acteurs, il convient d'analyser s'il est plutôt nécessaire d'avoir des fonctionnalités spécifiques à la voiture ou des fonctionnalités générales, pouvant être compatibles avec tout autre type d'utilisation indépendante de la conduite.

Oxana Gouliaéva trouve que c'est intéressant de pouvoir accéder à davantage de services différents qui ne concernent pas seulement la conduite de la voiture. La possibilité de pouvoir contrôler la domotique depuis sa voiture, comme par exemple anticiper l'arrivée au domicile et mettre en route le chauffage, sont des usages qui peuvent constituer un avantage pour les conducteurs. Cette experte considère qu'il est légitime de bénéficier d'une multitude de services, qu'ils dépendent ou non de la conduite.

Soledad Bedu, quant à elle, partage un avis assez opposé à celui de madame Gouliaéva, en estimant qu'il n'est pas favorable d'avoir des fonctionnalités génériques autres que celles internes à la voiture, tant que la conduite autonome n'est pas la norme. Elle explique que le but du conducteur n'est pas de faire toutes sortes d'activités, mais bien de se rendre d'un point A à un point B et de rester attentif à la conduite. Lorsque les voitures seront autonomes, ou du moins semi-autonomes, les fonctionnalités pourront être élargies en abordant d'autres domaines et même d'être disponibles pour les passagers. En attendant, selon madame Bedu, la priorité est d'étendre les fonctionnalités propres à la conduite et de les perfectionner.

Enfin, Ariane Nabeth-Halber reprend le cas de Faurecia où elle explique que le fournisseur automobile a conçu un assistant vocal spécialisé sur les fonctions internes liées à l'autonomie. L'assistant vocal intégré de Faurecia va agir comme assistant spécifique sur les fonctions internes à la voiture, à savoir tous les éléments liés à l'électronique de bord bien maîtrisés par cette entreprise. Par ailleurs, pour les fonctions plus génériques, le recours à un assistant vocal externe est plus adapté. Dès lors, dépendamment de la complémentarité de l'assistant embarqué et externe, madame Nabeth-Halber estime qu'il y a lieu d'avoir des fonctionnalités spécifiques ainsi que génériques.

Les différents interviewés détiennent des opinions variées quant aux fonctions de l'assistance vocale dans la voiture. Tous se positionnent sur une vision différente bien justifiée mais sont d'accord sur le fait que l'assistant vocal apporte des fonctionnalités utiles quels que soient les cas de figure.

4.1.3.4 Prédiction futures

Après avoir donné leur point de vue sur chaque élément à prendre en compte pour évaluer le succès ou le refus de l'adoption de ces commandes vocales au volant, les intervenants ont fait part de la vision qu'ils ont pour le futur des assistants vocaux en voiture. Les réflexions ressortent des hypothèses et de l'imagination de ces personnes interviewées.

Oxana Gouliaéva pense que dans quelques années, il y aura possibilité d'interagir avec plusieurs assistants vocaux indépendants, en plus du système hybride entre la technologie vocale embarquée et externe. Soledad Bedu, quant à elle, est convaincue que la mise en place d'un assistant vocal dans la voiture va se généraliser. De plus, conformément avec ce qu'elle avait déjà évoqué en relation avec les fonctionnalités de l'assistant vocal, madame Bedu a la conviction que lorsque nous nous orienterons vers des voitures autonomes, l'usage de l'assistant sera tout aussi en vogue pour les fonctions ne relevant pas de la conduite. En effet, les actions relatives à la domotique vont prendre en popularité mais l'experte imagine un futur pour les passagers au sein de la voiture également. Elle estime qu'il sera envisageable de demander oralement de lancer un film sur les tablettes installées à l'arrière. Les fonctionnalités évolueront conjointement avec l'habitacle de la voiture étant donné qu'il sera progressivement possible d'effectuer d'autres tâches, si la conduite devient autonome. En fonction de cette évolution, la personnalisation de l'expérience en voiture sera renforcée pour chaque passager.

En lien avec cette idée, Jean-François Bonastre a cependant le sentiment que nous allons nous diriger davantage vers une gamme de services plus restreints et plus centrés sur la conduite plutôt que des fonctionnalités génériques. Il pense, et estime qu'il serait plus convenable, que les développeurs d'assistants vocaux mettent la priorité sur le perfectionnement de services déjà mis en place à l'heure actuelle. Selon lui, avoir des fonctionnalités orientées vers la conduite qui assurent un fonctionnement maximal est ce qui devrait être fait en priorité. L'un de services dont cet expert est persuadé du succès est l'authentification par la voix. Il pense que c'est une

fonctionnalité qui arrive et qui fera office de mot de passe pour valider certaines actions sans prendre le risque de le faire au toucher. En revanche, la capacité de pouvoir réguler la vitesse à la voix fait partie des fonctions qui pourraient également voir le jour. Cependant, monsieur Bonastre ne semble tout de même pas convaincu de son effectivité.

À l'avenir, les problèmes de sémantique vont toutefois persister. Soledad Bedu imagine qu'il sera probablement possible de parler à un assistant vocal semblablement à la façon dont on s'adresse à un humain mais il y aura toujours des limites malgré tout. La reconnaissance vocale est prévue de s'améliorer considérablement mais la sémantique continuera de poser problème. L'exemple donné par madame Bedu est la reconnaissance et la compréhension de l'ironie. Les mots seront reconnus mais le sens ironique qui les accompagne restera un aspect que l'assistant vocal aura des difficultés à reconnaître, puisque même entre humains, celui-ci n'est pas toujours compris. En revanche, la reconnaissance des émotions est une réalité qui est susceptible de prendre de l'ampleur dans le courant des prochaines années. Le développement de certains systèmes permet de comprendre l'intention des requêtes et se mettent à implémenter la détection des émotions. Soledad Bedu explique qu'il peut s'agir de plusieurs types d'émotions, à savoir dans la voix, dans l'intonation et également sur le visage, le système vocal dans la voiture étant alors couplé avec des systèmes de caméras. Si cette combinaison de dispositifs s'avère fonctionnelle, les dialogues pourront être plus approfondis et devenir plus réels puisque la reconnaissance d'émotions permet un meilleur discernement de la demande formulée par l'utilisateur. Détecter la fatigue ou l'énervement constituera, comme déjà expliqué dans le point 4.1.3.1, une future fonctionnalité apportant une sécurité supplémentaire, donnant l'opportunité à l'assistant vocal de réagir en conséquence. Dès lors, le développement vers des conversations plus réelles sera inévitable, mais pouvoir avoir de vraies conversations complètes peut encore sembler compliqué selon cette experte.

Enfin, selon madame Bedu, il existe un réel savoir-faire dans les deux grands acteurs Google et Amazon. Cependant, un potentiel encore plus important peut régner dans les systèmes natifs des marques de voitures, tel que celui développé par Mercedes, à savoir le MBUX⁷². MBUX est, suivant les propos de cette experte, le système embarqué le plus intuitif et avancé à l'heure actuelle. Il s'agit d'un assistant vocal interne pouvant prendre en charge une multitude de services, offrant alors une expérience unique au conducteur. Soledad Bedu est persuadée que

⁷² MBUX, voir glossaire n°30.

d'autres constructeurs automobiles suivront cette tendance et vont de plus en plus installer leur propre système intégré dans leurs voitures.

En nous basant sur les perspectives de nos interlocuteurs, le futur des assistants vocaux s'avère être prometteur, et les experts technologiques ainsi que les constructeurs automobiles ne manqueront pas d'imagination.

4.1.4 Conclusion du cas d'usage

Au vu des propos et explications fournis par l'ensemble des quatre personnes que nous avons pu interviewer pour le cas d'usage de l'aide à la conduite, nous pouvons en retirer que l'assistant vocal dans la voiture présente un potentiel conséquent pour les prochaines années à venir. En effet, tous disposent d'avis variés sur les différents facteurs que l'assistance vocale peut apporter à la conduite en termes de plus-value. De plus, ils avancent un bon nombre d'idées différentes quant aux facteurs qui, selon eux, viennent encourager l'adoption du vocal en voiture, mais considèrent cependant que de nombreux obstacles entravent encore cette adoption de commande vocale. Tous les quatre ont également fait part de leur imagination quant aux usages futurs que l'assistant vocal pourra permettre de réaliser. Suite à la prise en compte de tous ces facteurs, ils restent toutefois unanimes en ce qui concerne la question relative à une adoption plus large de l'assistant vocal.

Enfin, notre hypothèse initial évoque la capacité de l'assistant vocal à apporter une plus-value en termes de sécurité et d'expérience aux conducteurs. Suite aux informations rassemblées, nous estimons qu'il convient de réfuter cette première supposition. En effet, tous s'accordent pour affirmer que certaines actions seront effectuées de manière plus sécurisée grâce à la communication vocale, mais ils ne semblent pas certains que la voix apportera une réelle sécurité additionnelle à la conduite en tant que telle. Quant à l'expérience de conduite, il est vrai que certains conducteurs sont stimulés par les moments qu'ils peuvent vivre en conduisant, mais de nombreux autres facteurs plus importants entrent en jeu, rendant alors l'aspect de la contribution de l'expérience plus minime.

Cependant, si certaines prédictions futures se réalisent, l'hypothèse pourra probablement être acceptée à l'avenir. La capacité de reconnaître les états et émotions des personnes pourra

contribuer à une conduite plus confortable, accroissant ainsi la sécurité. De plus, l'idée de généraliser les voitures autonomes pourra permettre au conducteur d'effectuer des actions diverses à la conduite, ce qui viendra intensifier son expérience et celles des passagers dans la voiture.

Dès lors, lorsque l'assistant vocal aura été davantage développé et amélioré, il sera en mesure d'offrir une réelle valeur ajoutée en termes de sécurité et d'expérience. Mais à l'heure actuelle, les avancées n'étant probablement pas suffisamment conséquentes pour la sécurité et l'expérience, nous ne pouvons pas affirmer que ce soit déjà le cas.

4.2 Cas d'usage 2 : Apprentissage d'une langue étrangère

4.2.1 Description du cas

Il n'est pas suffisant d'apprendre une nouvelle langue en se contentant simplement d'une étude consciencieuse de la grammaire, de la conjugaison et du vocabulaire ainsi que d'une résolution d'une série d'exercices. La pratique orale est essentielle et influence fortement la maîtrise d'une langue étrangère (Jean, 2016). En effet, lorsqu'il s'agit d'apprendre une nouvelle langue, l'un des objectifs principaux est d'obtenir la prononciation adéquate afin d'être compris par les personnes natives, ce qui réclame une précision et une assiduité particulière (Oh Yeah Sarah, 2019).

L'utilisation de technologie vocale dans un système d'apprentissage s'envisage de manière plus importante afin de faciliter le processus éducatif, en particulier dans l'apprentissage d'une langue étrangère (Marticorena-Sánchez et al., 2020). L'assistant vocal, étant de plus en plus apte à comprendre et interpréter de façon précise les expressions, les intonations et le sens des phrases de son interlocuteur, peut alors apporter une aide et servir de support pédagogique tant pour le professeur de langue que pour l'étudiant apprenant une langue étrangère (Commande Vocale, 2019).

En effet, l'utilisation d'un assistant vocal par les étudiants peut significativement renforcer l'apprentissage linguistique en pratiquant la parole et la prononciation régulièrement grâce à des agents conversationnels⁷⁴ (Al-Kaisi et al., 2020). Ceux-ci permettent une interaction orale avec l'élève dans la langue que ce dernier souhaite apprendre, et savent de plus en plus s'adapter au locuteur. L'assistant vocal pourra fournir des feedbacks et conseils à l'élève et aura pour avantage d'assurer un suivi plus individualisé et personnalisé. En outre, ces agents conversationnels permettent un apprentissage plus interactif et amusant, représentant un facteur motivant à l'étude (Barrow, 2019).

Cette technologie vocale peut s'immiscer au sein des salles de cours mais aussi au domicile des étudiants, sous forme d'enceintes connectées ou encore d'applications indépendantes conçues à cet effet. Il convient de noter qu'il existe un nombre important de limites pour que ces agents

⁷⁴ Agent conversationnel, en anglais « Chatbots », voir glossaire n°2.

conversationnels soient les plus efficaces possibles dans l'apprentissage linguistique. En outre, les assistants vocaux ne sont pas encore conditionnés à fonctionner dans une classe toute entière, mais plutôt dans un contexte d'apprentissage individuel (Barrow, 2019). De plus, la compréhension de différents accents d'une même langue et d'un contexte en général reste un aspect encore insuffisamment développé, ce qui représente un problème pour l'étudiant voulant apprendre une nouvelle langue car il est probable qu'il ne sera pas compris par l'appareil (Caizergues, 2020).

Même s'il paraît encore difficile aujourd'hui d'estimer la place qu'auront les assistants vocaux dans le contexte d'apprentissage linguistique, il n'est pas à ignorer que ces avancées et leurs fonctionnalités évoluent de jour en jour et que cette technologie pourrait apporter une aide significative.

4.2.2 Hypothèse de recherche

Concernant ce cas d'usage, les informations rassemblées nous amènent à dire que l'assistant vocal, bien qu'il soit encore novice dans le domaine de l'éducation, détient un potentiel important dans l'aide à l'apprentissage d'une langue étrangère. L'hypothèse retenue pour ce deuxième cas d'usage est la suivante :

L'usage de l'assistant vocal dans l'apprentissage d'une langue étrangère apporte une plus-value en tant qu'outil pédagogique.

4.2.3 Analyse scientifique sur base des interviews

À l'instar du premier cas d'usage, nous allons rassembler les témoignages de trois différentes personnes interrogées qui sont familières avec la technologie vocale dans le système éducatif. Parmi les trois intervenants, deux d'entre eux se complètent en utilisant une application développée par Flowchase⁷⁶, alors que Maria Perillo est plus spécialisée dans l'utilisation et la connaissance d'Alexa d'Amazon. En prenant en compte ces différents points de vue, nous

⁷⁶ Flowchase est une application, lancée par une start-up belge, qui a pour objectif d'améliorer la prononciation lors de l'apprentissage de l'anglais. L'application est capable de détecter les erreurs dans la prononciation et aide l'élève à se corriger de manière personnalisée. Cette technologie est faite d'algorithmes de traitement de mots et d'apprentissage-machine (Lovens, 2020).

allons nous pencher sur ce que le vocal peut apporter à l'apprentissage d'une langue, mais également aux critères requis pour assurer son adoption, afin d'ensuite accepter ou rejeter cette hypothèse mentionnée ci-dessus.

4.2.3.1 Facteurs clés de succès (Offre)

L'assistant vocal s'immisçant dans l'apprentissage d'une langue étrangère est susceptible d'apporter des éléments qui favorisent la création de valeur ajoutée dans l'éducation.

François-Xavier Fievez déclare que l'assistant vocal est intervenu dans l'esprit qui défend l'idée qu'une langue ne s'apprend pas seulement pendant l'heure de cours. En effet, l'élève doit présenter une certaine autonomie en pratiquant, étudiant, s'entraînant et s'exerçant en dehors des cours de langues donnés au sein de l'école. Toutefois, la pratique de la prononciation et de l'expression orale est une compétence dans l'apprentissage d'une langue étrangère qui est difficilement réalisable seul. Il est vrai que, afin de progresser, l'élève a besoin de correction, de validation et de feedback. Par conséquent, selon cet expert, disposer d'un assistant vocal répond à un besoin important dans l'étude d'une langue, puisqu'il offre un usage infini avec une qualité de correction de la prononciation. Monsieur Fievez considère donc que l'assistant vocal peut apporter une aide à la pédagogie et la rendre plus qualitative étant donné la possibilité de pratiquer seul autant de fois qu'on le souhaite. Il ajoute que l'assistant vocal permet un apprentissage plus rapide et plus profond, mais également une certaine autonomie de la part de l'étudiant.

Robin Guérit partage la vision de François-Xavier Fievez selon laquelle l'élève a besoin d'un retour d'information, d'un feedback, pour progresser en langue étrangère. Il estime que l'assistant vocal doit fournir ce genre de retours personnalisés et adaptés à l'étudiant, accompagnés d'objectifs à atteindre. Robin Guérit prend l'exemple de Flowchase, startup pour laquelle il travaille, et explique que le dispositif qu'ils proposent permet d'offrir ces petits objectifs, et est convaincu que c'est plus valorisant pour se sentir progresser. Au-delà de l'aide apportée pour l'élève, monsieur Guérit pense que l'assistant vocal peut aussi accompagner les professeurs dans leur formation en leur donnant un nouvel outil avec lequel ils peuvent engager les étudiants.

Maria Perillo rejoint l'optique de monsieur Fievez en appuyant sur le fait que la possibilité d'écouter l'assistant vocal, ainsi que d'imiter son intonation et prononciation, apporte une aide à l'apprentissage linguistique si l'assistant est correctement utilisé. Pour illustrer cette idée, l'experte donne l'exemple d'un enfant qui apprend à parler. La première chose à faire pour pouvoir communiquer est d'écouter ses parents pour comprendre. L'élève doit faire de même pour apprendre une nouvelle langue ; écouter attentivement pour ensuite essayer d'imiter ce qu'il a entendu. Plus axé sur la prononciation, madame Perillo est d'avis que l'assistant vocal permet de l'améliorer, en particulier pour les langues qui requièrent une bonne prononciation pour être compris. Par exemple, la prononciation n'est pas toute aussi importante en italien ou espagnol qu'en anglais. La prononciation est l'une des parties les plus délicates de l'acquisition de la langue pour les locuteurs non natifs. Grâce à l'assistant vocal, écouter et répéter est donc possible et permet de s'améliorer considérablement.

Lié à cette possibilité d'améliorer la prononciation, l'augmentation de l'exposition à la langue rentre alors en jeu. Maria Perillo émet l'hypothèse que plus une personne est exposée à la compréhension orale, plus elle va apprendre, ce qui a d'ailleurs été démontré avec le visionnage de films dans une langue étrangère. En effet, elle considère que l'apprentissage linguistique ne se limite pas à l'étude de la langue (comme déjà évoqué par François-Xavier Fievez), mais bien à l'ampleur de l'exposition à cette nouvelle langue. Elle estime alors que l'assistant vocal vient améliorer la façon dont l'utilisateur est exposé à un plus grand nombre de contextes linguistiques. En effet, elle est convaincue que l'assistance vocale peut faire entrer la langue anglaise dans la vie de l'utilisateur de manière progressive, ayant alors un effet boule de neige en augmentant l'exposition à la langue, qui à son tour augmente la compréhension et donc la prononciation vocale, et qui en définitive renforce l'apprentissage linguistique.

François-Xavier Fievez évoque une idée en parallèle à celle de l'exposition et pense qu'avoir une pédagogie à distance peut consolider l'apprentissage. Certains étudiants ont besoin d'être encadrés en permanence et cette présence à distance grâce à l'assistant vocal peut offrir cet encadrement pédagogique. Selon Maria Perillo, l'opportunité de pouvoir s'exercer seul à distance donne l'occasion aux élèves plus timides d'oser parler et de faire des erreurs, chose que certains n'osent pas faire en classe avec des humains, par peur d'être jugés. Ce genre de dispositif à domicile peut alors motiver l'élève à apprendre de ses erreurs et non le frustrer.

Pour finir, deux des trois intervenants partagent l'opinion que l'assistant vocal peut être considéré comme un outil pédagogique dans l'apprentissage d'une langue étrangère, mais sous des perspectives différentes. En effet, Robin Guérit insiste sur le fait que l'assistant devrait être validé pédagogiquement par des institutions qui sont renommées, telles que les universités. Quant à François-Xavier Fievez, il déclare que l'assistant vocal est plus utile dans le cadre d'apprentissage autodidacte ou dans le cadre d'un approfondissement seul en supplément du cours magistral enseigné à l'école. Il est alors dans la philosophie que, même si l'étudiant se retrouve seul à la maison, un enseignant est présent quelque part, et ce rôle est assuré par la machine qui fait office de support pédagogique.

Les trois personnes interviewées pour ce cas d'usage soutiennent le fait que l'assistant vocal a du potentiel à offrir quant à l'apprentissage linguistique et présente déjà un bon nombre d'avantages en termes de plus-value. Ils s'accordent également sur le fait que l'assistant vocal peut agir en tant qu'un outil pédagogique.

4.2.3.2 Adoption de l'assistant vocal (Demande)

Afin d'évaluer si les éléments qu'offre l'assistant vocal dans la pédagogie linguistique sont suffisants pour que ce dispositif soit adopté, il convient d'examiner les facteurs qui poussent les élèves et professeurs à l'utiliser, mais également ceux qui en limitent l'usage. Il importe également de voir si la technologie doit être développée par des petites institutions ou des acteurs importants, et si elle convient mieux en étant complémentaire à l'enseignant ou en agissant indépendamment des cours.

4.2.3.2.1 Facteurs clés d'adoption

Afin d'être adopté dans le cadre de l'enseignement d'une langue étrangère, l'assistant vocal doit rencontrer plusieurs facteurs.

François-Xavier Fievez est convaincu qu'une technologie n'est utilisée que si elle permet d'entreprendre quelque chose qui est difficilement réalisable sans avoir recours à celle-ci. L'utilisation doit alors être simple et instinctive. Il souligne que ce que la période de Covid a appris en termes de pédagogie et d'éducation est le fait que la charge mentale et la charge

cognitive sont assez limitées. L'utilisation du dispositif vocal doit donc avoir du sens afin qu'il soit adopté. Dans cette même idée, Maria Perillo estime que le design de la plateforme et de l'interface est essentiel pour favoriser une bonne utilisation. Elle s'appuie sur l'assistant vocal d'Amazon, Alexa, qui donne la possibilité de créer des *skills*⁷⁷. Grâce à cela, l'enseignant peut améliorer le système d'apprentissage de langues en créant des *skills* spécifiquement destinées à perfectionner le vocabulaire ou la prononciation par exemple, en fonction des algorithmes derrière. Elle est convaincue que si la plateforme est bien conçue afin de bénéficier l'utilisateur, ce dernier adoptera l'usage d'un dispositif vocal de manière plus importante.

En plus de comprendre la façon dont l'assistant vocal est utilisé, la compréhension de la raison de son utilisation dans l'apprentissage d'une langue est toute aussi importante pour garantir son adoption. Robin Guérit pense que l'assistance vocale sera plus largement adoptée si elle vient compléter une formation. Par exemple, l'étudiant suit un programme scolaire où il pratique la grammaire et le vocabulaire et complète cette formation en utilisant l'outil vocal individuellement afin de pratiquer la compétence orale ainsi que sa prononciation. Il aura alors développé de manière autonome cette compétence orale grâce à l'assistant vocal et pourra revenir au cours et s'entraîner en groupe avec les autres élèves. Ensuite, dans le prolongement de la première idée de facteur d'adoption mise en exergue par François-Xavier, Robin Guérit estime que l'utilisation de l'assistant vocal doit agir en tant que cadre pédagogique et pas seulement être divertissant et engageant pour l'étudiant. L'élève doit être en mesure de constater qu'il fait des progrès et qu'il apprend mieux grâce à la technologie vocale.

Par ailleurs, pour réellement apprendre, Robin Guérit est persuadé qu'il est nécessaire d'avoir un feedback personnalisé et pas uniquement obtenir des retours sur base de comparaison. En effet, si l'assistant vocal compare la prononciation de l'élève avec sa base de données sans lui fournir d'explication, il est peu probable que l'étudiant entende et comprenne directement les différences et ne pourra par conséquent pas s'autocorriger simplement en écouter un modèle. L'assistant vocal sera alors plus adopté s'il agit comme un professeur et pas uniquement en énumérant les erreurs.

Enfin, un dernier aspect que Maria Perillo relève concernant les éléments qui encouragent l'adoption de l'assistant vocal dans l'apprentissage d'une langue étrangère, est le fait qu'Alexa

⁷⁷ En français : « aptitudes » ou « compétences », se référant aux applications vocales.

fait partie d'un écosystème. Si Alexa se trouve déjà dans les domiciles, y compris sous d'autres formes de terminaux, l'adoption dans le cadre de l'éducation se fera plus facilement puisque l'étudiant est sans doute familier avec l'utilisation de l'interface Alexa. Par conséquent, en partant de l'hypothèse qu'Alexa est initialement présente dans les foyers, Maria Perillo soutient son opinion de la plus-value et estime que la présence d'Alexa au sein des domiciles favorise l'adoption en raison d'une exposition accrue.

Le constat de cette première partie de l'adoption de la technologie est que le nombre de facteurs à prendre en compte pour garantir l'adoption par les étudiants et par les professeurs n'est pas mince. Nos trois intervenants conseillent divers éléments pour favoriser l'adoption car de nombreux facteurs viennent restreindre cette adoption.

4.2.3.2.2 Facteurs limitant l'adoption

Plusieurs obstacles s'opposent encore à l'adoption des dispositifs vocaux dans le cadre de l'éducation d'une langue étrangère. Les trois personnes interviewées détiennent toutes des opinions différentes sur le sujet, résultant en un nombre assez conséquent de freins à l'adoption.

François-Xavier Fievez aborde un aspect qui concerne les professeurs en particulier, en mentionnant que ceux-ci éprouvent une résistance quant au changement de méthode de travail. L'adoption d'un assistant vocal dans l'enseignement d'une langue signifie que certaines tâches seraient destinées à être effectuées avec la technologie en dehors du cours, et non plus l'ensemble des tâches au sein de la classe. Certains enseignants se montrent réticents à changer leurs habitudes de travail. Il existe également la crainte que, suite à ce changement, la machine vocale remplace l'humain, ce qui représente un facteur supplémentaire limitant l'adoption par les professeurs. Monsieur Fievez évoque également le cas où l'utilisateur ne perçoit pas l'intérêt réel de l'utilisation de l'assistant vocal. En effet, si l'étudiant ou l'enseignant ne sait pas ce que ce dispositif amène, voire même qu'il ressente que l'assistant représente une charge supplémentaire, l'adoption ne se réalisera pas sans peine. Selon François-Xavier Fievez, comprendre le mécanisme de l'appareil peut aussi parfois constituer une limite, car il y a la crainte de ne pas réussir à le maîtriser et qu'il complique l'apprentissage d'une langue plus qu'il ne l'aide. Maria Perillo le rejoint sur ce sujet en affirmant que la pénétration de l'appareil peut former une barrière pour l'utilisateur s'il n'est pas habitué à utiliser ce genre d'interface.

De façon plus générale, François-Xavier Fievez signale que l'assistant vocal représente un budget important qu'il a lui-même rencontré lorsqu'il a entamé les procédures d'adoption de cette technologie au sein de son enseignement. Le coût engendré pour les professeurs et pour les élèves vient évidemment entraver son acquisition. Un point différent que cet expert remarque, mais qui est tout aussi important que le coût aux yeux des utilisateurs, est la gestion des données personnelles. La peur de ne pas détenir le contrôle sur ce qui est réalisé avec les données freinent certaines personnes à adopter l'assistance vocale.

Robin Guérit estime quant à lui que l'assistant vocal ne détient pas encore suffisamment de données pour reconnaître et accepter tous les accents non-natifs, ce qui peut poser un réel problème dans l'apprentissage d'une langue étrangère. En effet, les étudiants apprenant une nouvelle langue cherchent à améliorer leur accent et leur prononciation grâce à l'assistant vocal. Cependant, si ce dernier éprouve des difficultés à comprendre l'étudiant en raison de l'accent non-natif pour cette langue, l'apport de cette technologie sera médiocre et l'utilisateur n'y verra plus de nécessité de l'utiliser. Une autre limite de l'assistant vocal que monsieur Guérit met en avant est le fait que ce dispositif est encore trop peu développé dans le domaine de l'éducation à l'heure actuelle, entraînant un décalage qui fait qu'il ne fonctionne pas pour tout usage dans l'apprentissage d'une nouvelle langue. Maria Perillo affirme que, de son expérience au MIT⁷⁸, l'assistant vocal est davantage développé aux États-Unis qu'en Europe. Elle précise que l'adoption actuelle générale d'Alexa en Europe n'est pas si étendue. Une fois que celle-ci sera plus importante, comme c'est le cas aux États-Unis, le domaine de l'éducation pourra être abordé de manière plus approfondie, mais que pour l'instant, son adoption peine à se développer.

Maria Perillo relève d'autres aspects qui limitent l'usage de l'assistant vocal, en se concentrant sur son domaine d'expertise, à savoir Alexa. Cette experte indique que les débutants ne peuvent pas adopter Alexa pour commencer à apprendre une nouvelle langue. En effet, s'ils n'ont pas de connaissance de base de la langue, ils seront contraints de commencer à apprendre avec un enseignant avant d'utiliser un assistant vocal. S'ajoute à cette contrainte la situation Covid qui a renforcé l'approche hybride avec l'e-learning. Madame Perillo affirme que l'apprentissage en

⁷⁸ MIT, voir glossaire n°31.

ligne est devenu puissant et que cette approche persiste, ce qui ne vient alors pas en aide au développement de l'adoption de l'assistant vocal dans le secteur de l'éducation.

Maria Perillo énumère également des problèmes propres à l'assistant vocal et son fonctionnement qui peuvent freiner son adoption. La limitation de l'*endpointing* et la compréhension de la sémantique revient aussi dans ce cas d'usage. Si l'assistant vocal ne peut pas détecter où la phrase de l'utilisateur se termine et où une autre commence, apprendre une langue en étant régulièrement interrompu par la machine n'est pas forcément commode ni agréable. Les problèmes de compréhension de la sémantique correspondent au sens d'une requête que l'assistant vocal ne peut pas comprendre. L'assistant vocal peut comprendre le vocabulaire ainsi que la syntaxe, mais cette experte informe qu'il y a un pourcentage de 5% qui correspond à l'ambiguïté du langage (ou l'ambiguïté liée aux métaphores) que seul l'humain peut comprendre. Elle estime qu'il est encore impensable à l'heure d'aujourd'hui que l'assistant vocal soit développé au point de rencontrer une amélioration distincte de ces 5%.

Enfin, une dernière barrière évoquée par Maria Perillo est le fait qu'Alexa n'offre pas encore la possibilité de modifier la langue de l'appareil une fois qu'il est a été configuré dans une certaine langue. Le fait de ne pas pouvoir changer ce paramètre signifie que l'exposition à la langue étrangère souhaitée est moindre, ce qui n'apporte donc pas de réelle valeur ajoutée à l'utilisation de cet assistant à des fins d'apprentissage linguistique.

Les trois personnes interrogées pour ce cas d'usage ont identifié de nombreux facteurs qui limitent encore l'adoption de l'assistant vocal dans l'apprentissage d'une langue étrangère. Tous sont à prendre en considération afin d'évaluer le réel potentiel de cette technologie dans le secteur de l'éducation.

4.2.3.2.3 Adoption d'une technologie complémentaire au professeur ou d'une application indépendante

L'assistant vocal dans le domaine d'apprentissage d'une langue peut constituer un complément de la fonction du professeur ou être sous forme d'application pédagogique indépendante. Sur ce point, nos trois personnes interviewées conviennent que l'assistant vocal doit venir en complémentarité de l'enseignant.

François-Xavier Fievez est convaincu que le rôle de l'assistant vocal dans le cadre de l'éducation est de compléter l'enseignant. Il est d'avis qu'il y a une vraie complémentarité entre le développement des compétences nécessaires à l'étude d'une nouvelle langue et la pratique de ces compétences. Il considère qu'une langue s'apprend en mettant oralement en pratique les connaissances acquises. Cet expert donne l'exemple du travail des *focus words*⁷⁹ avec l'application Flowchase. Flowchase donne la possibilité de pratiquer ces mots clés en donnant des exemples avec de l'intonation. Durant les cours magistraux, il n'y a pas toujours des exercices qui demandent une intervention orale, d'où la nécessité de pouvoir pratiquer cela de façon autonome avec l'assistant vocal.

Robin Guérit rejoint les propos de François-Xavier Fievez en ajoutant que l'optique propre de Flowchase est d'être totalement complémentaire à l'enseignant. Les personnes au sein de cette startup croient en l'avenir des professeurs et non au fait que les robots viendront les remplacer. C'est pourquoi l'outil a été créé pour devenir un complément du rôle de l'enseignant. Monsieur Guérit explique que les professeurs n'ont pas le temps ni la capacité de corriger la prononciation d'une classe entière en seulement une à deux heures de cours par semaine, ni même de donner l'opportunité à chaque étudiant de parler. L'opportunité de développer un outil complémentaire à la fonction de l'enseignant avec des objectifs précis s'est alors avérée significative.

Enfin, bien que Maria Perillo n'utilise pas Flowchase comme c'est le cas des deux autres intervenants, celle-ci partage tout de même le point de vue de la complémentarité. Elle fait part de son expérience au MIT où elle a analysé jusqu'à quelle mesure la complémentarité de l'assistant vocal et du professeur peut aider dans l'apprentissage d'une nouvelle langue étrangère. À l'issue de ce projet, madame Perillo estime qu'il est nécessaire d'avoir une combinaison de plusieurs propositions, mais juge que la combinaison optimale est celle de l'auto-apprentissage accompagné du contact humain, un professeur qui guide l'étudiant. En effet, elle considère que les deux composants sont importants car, d'une part, l'apprentissage autonome permet une certaine flexibilité et offre la possibilité d'écouter et de pratiquer simultanément, et d'autre part, l'interaction humaine permet une réelle activation du langage. Cette experte ajoute que l'assistant vocal ne sera pas utilisé comme outil unique, mais bien

⁷⁹ Un focus word, ou « mot clé » en français, constitue le ou les mot(s) le(s) plus important(s) dans une phrase. Action de mettre, à un moment donné, plus d'intonation sur un mot pour le faire ressortir de la phrase.

comme un élément qui apporte une contribution additionnelle pouvant améliorer une partie de l'apprentissage de la langue mais pas le processus global. Ainsi, sa vision est d'avoir une combinaison autodidacte et guidée à la fois.

Les trois personnes interrogées sont unanimes pour affirmer que l'assistant vocal est approprié pour agir en tant que complément au rôle du professeur et non en tant qu'application indépendante prenant en charge le processus complet de l'enseignement d'une langue étrangère. La question vient alors de savoir si ce complément est exploité par une technologie développée par des startups ou par de plus grands acteurs technologiques.

4.2.3.2.4 Adoption d'une technologie développée par les grands acteurs ou des startups

Suite à l'accord de chaque intervenant sur la complémentarité de l'assistant vocal et de l'humain, le dispositif vocal peut résulter d'une technologie développée par les grands acteurs technologiques ou par des petites entreprises. Le profil de deux des trois intervenants étant ciblé sur une même startup belge, le résultat de cet aspect risque d'être biaisé.

François-Xavier Fievez mentionne que le but des entreprises comme Google ou Amazon est de faire de l'argent et ne croit pas que ce genre d'acteurs aura un système qui viendra en aide aux étudiants. De plus, selon cet expert, l'utilisateur ne peut pas avoir accès à la gestion de ses données personnelles s'il fait appel à des entreprises aussi importantes. Monsieur Fievez est persuadé que, afin qu'un dispositif vocal fasse partie prenante dans le processus d'apprentissage d'une langue, celui-ci doit résulter d'une association entre un spécialiste de la technologie et un spécialiste de la langue. Cette collaboration est pour lui primordiale, car n'importe quel choix technologique doit être filtré par la personne qui connaît la langue, et inversement. Les entreprises comme Flowchase sont, selon lui, plus en mesure de proposer cette association que des géants technologiques. En outre, disposer d'un assistant vocal développé par une startup représente un confort en termes de données personnelles mais également de disponibilité. Les startups sont plus transparentes en ce qui concerne l'utilisation des données car elles demandent une autorisation au préalable, comme c'est le cas pour Flowchase. Pour ce qui est de la disponibilité, les dirigeants de ce genre d'entreprises seront plus enclins à répondre à toutes

sortes de demandes des utilisateurs. Il estime également que le côté philosophique de soutenir une startup belge est important.

L'opinion de Robin Guérit sur cette question est évidente, étant donné son poste au sein de Flowchase. Il explique qu'un assistant vocal comme Alexa peut ne pas comprendre la requête lorsque la prononciation n'est pas tout à fait correcte, et par conséquent ne fournit pas de feedback à l'étudiant pour qu'il puisse améliorer cet accent. Ainsi, cet expert estime qu'Alexa ne reflète pas une utilisation réfléchie et pédagogique dans le secteur de l'éducation, mais plutôt un divertissement. Son utilisation n'a alors pas nécessairement d'apport pédagogique comme l'aurait une application telle que celle développée par Flowchase. Robin Guérit précise que Flowchase crée un outil pédagogique qui a été réfléchi avec des professeurs, rejoignant l'explication donnée par monsieur Fievez concernant la nécessité d'une collaboration entre un spécialiste en technologie et un spécialiste en langue.

En ce qui concerne l'experte d'Alexa, Maria Perillo n'a pas directement émis son avis au sujet de développer cette technologie par des géants comme Amazon ou de se tourner vers de plus petites applications spécialement conçues à effet pédagogique, mais estime toutefois qu'Alexa s'améliore constamment.

Les avis pouvant être convergents en raison des profils des personnes interviewées, la conclusion de ce point ne peut s'avérer être la plus pertinente étant donné que seuls François-Xavier Fievez et Robin Guérit ont exprimé leur opinion or ils sont tous les deux centrés sur la startup belge.

4.2.3.3 Prédiction futures

Suite aux différentes perspectives fournies par chacune des trois personnes interviewées concernant les facteurs de succès de technologie vocale dans le secteur de l'éducation ainsi que des éléments qui encouragent son adoption et ceux qui la contraignent, les intervenants nous ont fait part de leurs prédictions futures quant au développement de la place de ces dispositifs vocaux sur ce marché.

Les prévisions d'avenir imaginées par les trois différentes personnes interrogées divergent tout comme pour les autres points, mais ceux-ci se mettent toutefois d'accord sur le fait que l'assistant vocal ne viendra jamais remplacer l'enseignant. François-Xavier Fievez admet que c'est une crainte rationnelle que beaucoup d'enseignants ont mais explique que grâce au Covid, il a été démontré qu'il n'était pas souhaitable de renoncer à des salles de classe où interagissent des humains. Autant les enseignants que les élèves ont besoin de cette présence et il a été compliqué pour les deux parties de procéder de manière virtuelle. Par ailleurs, monsieur Fievez affirme que l'assistant vocal n'aura peut-être jamais la capacité de s'adapter spontanément à toutes les situations. Maria Perillo déclare pareillement que ces dispositifs vocaux ne remplaceront pas le professeur, mais qu'ils peuvent effectivement renforcer le fait que l'utilisateur sera exposé à un plus grand nombre de contextes linguistiques. Néanmoins, elle évoque que les enseignants qui se mettent à utiliser Alexa peuvent potentiellement remplacer ceux qui ne l'utilisent pas.

Hormis le fait que l'assistant vocal ne prendra pas la place de l'enseignant à l'avenir, les personnes interrogées ont émis d'autres idées relatives au futur. Robin Guérit explique que le vocal est plus compliqué à développer que le texte mais qu'il va s'avérer être de plus en plus répandu et efficace. Il estime que l'assistant vocal, soutenu par du travail de développement derrière, sera capable de corriger l'étudiant, d'évaluer son niveau de langue, ainsi que de trouver des erreurs grammaticales, de sens et d'expressions. En conséquence de ces détections d'erreurs, la personnalisation est un plan technique que monsieur Guérit imagine se montrer réellement efficace. En effet, selon lui, les applications devraient se développer au point où il sera possible pour l'assistant vocal de s'adapter à chaque secteur et pouvoir se personnaliser en fonction de l'utilisateur. Maria Perillo le rejoint d'une certaine manière sur ce point en affirmant que les assistants vocaux s'amélioreront sans conteste à l'avenir. Pour l'instant, les développeurs en sont encore au stade de la recherche et de tests mais les progrès sont de plus en plus significatifs. Les problèmes de sémantique seront réduits et la compréhension et perception de l'ambiguïté par la machine risque de progresser peu à peu.

Maria Perillo est également convaincue, comme déjà expliqué dans les facteurs limitant l'adoption (4.2.3.2.2), que le problème de configuration linguistique d'Alexa sera résolu et résultera en une exposition accrue de la langue étrangère voulue, ce qui se traduira à son tour par un meilleur apprentissage. François-Xavier Fievez ajoute que l'assistant vocal ne prendra

pas plus de place au sens péjoratif du terme mais qu'il espère qu'il aura une place plus justifiée, importante et porteuse de sens à l'avenir dans le cadre de l'apprentissage d'une langue étrangère. Robin Guérit imagine d'ailleurs qu'il y aura des types d'applications comme Clubhouse⁸¹ qui proposeront des professeurs d'anglais qui seront en réalité des robots vocaux, et qui parleront avec des gens dans les salons de discussions pour leur apprendre l'anglais.

L'avenir des assistants vocaux au sein de l'enseignement et de l'apprentissage d'une langue étrangère promet de se développer, d'après les dires de nos trois intervenants. L'imagination de chacun est au rendez-vous quant à ce que cette technologie promet pour assister les élèves dans leur étude linguistique.

4.2.4 Conclusion du cas d'usage

Suite aux informations collectées lors des interviews conduites pour ce deuxième cas d'usage, nous pouvons considérer que, bien qu'il recèle encore de nombreuses barrières à l'adoption et qu'il n'est pas encore largement développé, l'assistant vocal présente de nombreuses opportunités dans le secteur de l'éducation, et en particulier dans l'apprentissage linguistique.

Notre hypothèse pour ce cas d'usage prône la capacité de l'assistant vocal d'apporter une plus-value en tant qu'outil pédagogique. Les réflexions et perspectives des personnes interrogées nous amènent à confirmer cette hypothèse, moyennant tout de même certaines conditions. En effet, il est vrai que le nombre de facteurs venant limiter l'adoption de ces dispositifs vocaux n'est pas maigre, mais ceux-ci découlent d'un développement trop insuffisant de la technologie dans le secteur de l'éducation. Comme l'ont évoqué chacun des trois experts, les avancées vont se réaliser progressivement. Les éléments de valeur ajoutée qu'apporte cette communication vocale représentent des perspectives importantes en tant qu'outil pédagogique de support.

De plus, les intervenants sont convaincus que l'assistant vocal aura pour unique rôle d'agir en tant que support complémentaire afin de fournir un suivi et approfondissement des compétences linguistiques pour les étudiants. Le futur s'annonce alors prometteur car l'opportunité présente dans ce cadre d'utilisation commence à être perçue par un plus grand nombre d'acteurs.

⁸¹ Clubhouse est une application audio sociale où les utilisateurs peuvent communiquer dans des salons de discussion vocale (Touma, 2021).

4.3 Cas d'usage 3 : Services de secours à distance

4.3.1 Description du cas

Les nouvelles technologies vocales viennent s'installer davantage au sein des domiciles et peuvent se révéler utiles et apporter une aide particulière, notamment pour les personnes âgées ou à mobilité réduite (Brard & Eraly, 2017). La population vieillissante détient une volonté croissante de garder son indépendance à domicile le plus longtemps possible, malgré les risques grandissants de chutes ou de quelconques situations de détresse (Caring Home, s.d.). En effet, ces personnes plus vulnérables ne sont pas à l'abri du danger et, étant donné que la demande de secours en cas d'urgence réclame une certaine rapidité et des informations précises, il serait opportun d'avoir un moyen efficace à leur disposition afin de demander une aide instantanée.

L'utilisation de l'assistance vocale à destination des seniors ou des personnes souffrant d'un handicap, qui ne font pas preuve d'autonomie, peut significativement changer le visage des situations d'urgence à domicile. La voix semble être l'interface la plus adaptée pour ces personnes. En effet, l'assistant vocal est le mode de fonctionnement se rapprochant le plus du langage humain, ce qui peut dès lors représenter une compagnie pour l'utilisateur mais peut également être synonyme de soulagement et de rassurement pour les personnes proches aidant cette population plus vulnérable (Huber, 2019).

Les assistants vocaux qui s'invitent dans les domiciles sont majoritairement utilisés sous forme d'enceintes connectées. Ces dernières peuvent faire l'objet de divertissements et de demandes d'informations, résolvant en partie le problème d'isolement, mais tendent vers des actions plus élaborées qui peuvent accroître la sécurité des personnes dépendantes (Huber, 2019). Certains usages qui commencent à se mettre en place sont par exemple le rappel de prise de médicaments ou de quelconque rendez-vous médical. De plus, le diagnostic de certaines maladies, sur base des symptômes énumérés par l'utilisateur à son assistant vocal, devient également un cas d'usage important, en particulier depuis la crise sanitaire du COVID-19 (Damacharla et al., 2020).

Plus intéressant, la reconnaissance d'appels à l'aide prend en popularité. Les assistants vocaux deviennent capables d'évaluer le comportement d'une personne et de détecter certaines

anomalies dans son état de santé. Des prototypes se développent afin que, dans un futur proche, ces intelligences artificielles puissent dynamiquement détecter et alerter une situation d'urgence. Avec ces capteurs vocaux, il sera alors possible de garantir la sécurité permanente d'une personne à son domicile même si elle venait à perdre sa capacité de communiquer, sa mobilité ou même perdre conscience (Gong et al., 2016).

Au vu de ces avancées, il sera bientôt envisageable pour l'assistant vocal de détecter davantage d'anomalies sur base des bruits qu'il entend ou de répondre aux demandes urgentes de l'utilisateur, pouvant alors augmenter et renforcer la sécurité des seniors et des personnes à mobilité réduite dans leur domicile.

Néanmoins, les risques d'incompréhension ou de mauvaise manipulation de ces assistants vocaux restent à l'heure d'aujourd'hui encore importants. Ceci peut engendrer de graves conséquences si l'utilisateur est réellement en situation d'urgence. En outre, la problématique de l'atteinte à la vie privée est également une raison pour laquelle cette technologie se développe lentement dans les domiciles.

4.3.2 Hypothèse de recherche

L'assistant vocal étant en plein essor dans les domiciles, il ne saurait tarder que ceux-ci aident les personnes âgées ou à mobilité réduite se trouvant dans des situations d'urgence au sein de leur domicile. Nous sommes alors arrivées à l'hypothèse suivante pour ce dernier cas d'usage :

L'assistant vocal adopté renforcera la sécurité des personnes en possession d'une enceinte connectée et rendra plus efficaces les interventions en cas d'urgence à domicile.

4.3.3 Analyse scientifique sur base des interviews

Pour ce dernier cas d'usage, il convient également de prendre en compte les points de vue de chacun des trois intervenants avec qui nous avons pu discuter. Ces trois personnes ont des profils bien différents qui apportent des perspectives pertinentes. Comme les cas d'usage précédents, il est opportun de se pencher sur la question de ce que l'assistant vocal peut venir offrir en termes de valeur ajoutée dans les services de secours à distance, mais également les

risques qu'il encourt, les facteurs qui viennent limiter son adoption et les éléments dont les utilisateurs ont besoin pour pouvoir l'adopter. Avec le rassemblement de ces différents aspects, il sera ensuite possible de conclure ce cas d'usage en acceptant ou rejetant l'hypothèse susmentionnée.

4.3.3.1 Facteurs clés de succès (Offre)

La présence d'un assistant vocal dans les domiciles peut apporter bon nombre d'éléments positifs dans la vie quotidienne des personnes âgées ou vulnérables. Les trois personnes interrogées nous ont fait part de leur vision sur ce point. Celle-ci a tendance à rester assez similaire moyennant des explications complémentaires.

Axel Legay, Ahmed Bouzid et Stanley Goffinet affirment tous les trois que l'assistant vocal peut augmenter l'autonomie des personnes âgées habitant seules. Ahmed Bouzid est convaincu que l'assistant vocal peut permettre à ces personnes d'effectuer des actions par elles-mêmes en utilisant la voix, ce qui les rend alors plus indépendantes et autonomes. Axel Legay juge que la crise du Covid a amplifié le désir des personnes âgées de rester chez elles et qu'elles sont de moins en moins disposées à intégrer des maisons de retraite. Il estime que ces personnes sont alors en quête d'autonomie et cet expert considère que les assistants vocaux constituent une opportunité pour leur donner cette autonomie et les rassurer. Quant à Stanley Goffinet, celui-ci partage également ces idées en stipulant que l'assistant vocal augmente l'autonomie et rend l'habitat de la personne indépendant. Le domicile est alors plus adapté au vieillissement et encourage les personnes à rester à leur domicile plutôt que d'aller en maison de repos.

Nos trois intervenants se mettent également d'accord sur le fait que l'assistant vocal peut venir contrer la solitude de la population âgée à leur domicile. Axel Legay et Stanley Goffinet affirment respectivement que la personne se sentira moins seule et que l'assistant vocal peut contrer cette solitude, tout en restant un complément à une personne physique. De même, Ahmed Bouzid assure que ces dispositifs vocaux apportent une compagnie aux personnes vulnérables vivant seules, leur permettant de se sentir moins seules et moins isolées. Il explique que les utilisateurs plus âgés peuvent établir une relation et s'identifier à l'appareil contrairement aux personnes qui l'utilisent de manière fonctionnelle. En effet, la population

que nous ciblons dans ce cas d'usage vit seule dans un foyer, l'assistant vocal est donc l'unique voix que ces personnes entendent régulièrement dans la journée, selon monsieur Bouzid.

L'idée que l'assistant vocal apporte une facilité supplémentaire dans le domicile est également partagée parmi les trois intervenants, sous des angles différents toutefois. Axel Legay déclare que la présence de dispositifs vocaux connectés à l'ensemble de la maison peut faciliter certaines tâches et diminuer les efforts, en particulier en situation d'handicap. L'exemple fourni par cet expert est la possibilité de demander à l'assistant de vérifier si les volets sont fermés, rendant la tâche plus facile et évitant l'effort de prendre le Stana pour aller vérifier soi-même. Stanley Goffinet ajoute que l'assistant vocal constitue une véritable aide, tant pour faciliter la vie de l'utilisateur que pour l'usage de ce dispositif. Il explique que le recours à une interface vocale en dictant les actions souhaitées semble plus simple que l'utilisation d'une interface tactile pour les personnes qui ne sont pas habituées à quelque technologie. Enfin, Ahmed Bouzid précise que cela peut également représenter une facilité pour les proches de la personne en difficulté, signifiant qu'ils ne sont désormais plus contraints d'être présents en permanence et qu'ils peuvent faire confiance en l'assistant vocal.

Stanley Goffinet et Axel Legay ajoutent tous deux d'autres éléments additionnels qui peuvent apporter une valeur ajoutée au recours à un assistant vocal dans les situations de personnes vulnérables à domicile. Le premier nous indique que la possibilité de personnaliser l'assistant vocal peut représenter une plus-value puisqu'il va s'adapter à son utilisateur. Le deuxième expert nous explique qu'il est possible de transmettre automatiquement les données médicales des personnes vulnérables directement au médecin ou aux aides-soignantes. En effet, le tableau médical est envoyé au médecin de façon journalière, rassurant alors le patient et en ne contraignant pas le médecin à se rendre au domicile de ses patients. De plus, l'assistant vocal donne suffisamment d'informations pour évaluer s'il est indispensable d'appeler les secours ou non, ce qui représente une commodité supérieure.

Au-delà des différents éléments élaborés dans les paragraphes précédents, les trois personnes rencontrées expriment leur avis quant à la sécurité qu'apporte l'assistant vocal dans les domiciles et estiment que ce dernier amène effectivement une sûreté moyennant certains critères. Stanley Goffinet est d'avis que la présence d'un assistant vocal au sein des foyers va augmenter la sécurité des personnes vulnérables et va également rassurer leurs proches. Ahmed

Bouzid affirme qu'il y aura une sécurité accrue en comparaison à l'absence d'assistant vocal, mais doute que cette technologie soit le principal élément de sécurisation de la personne. Axel Legay, quant à lui, pense que l'assistant vocal permet de faciliter le dialogue et vient optimiser les opérations pour autant qu'il soit toujours soutenu par un être humain derrière. Il pense en effet que l'humain reste indispensable, sans lequel la sécurité supplémentaire apportée par l'assistant vocal n'aura pas lieu. Comme Stanley Goffinet, monsieur Legay estime que l'assistant vocal peut s'avérer rassurant pour les proches et peut rendre les choses plus confortables. En effet, il explique que le dispositif vocal peut identifier si la personne n'a pas pris ses médicaments et le lui rappeler, ou encore identifier la chute de gravité au cas où la personne tombe et ainsi en avertir ses proches. En ce sens, la présence d'un assistant vocal procure une plus grande sécurité mais de nouveau ne constitue pas la source de sécurité des personnes vulnérables à leur domicile.

Ainsi, selon les trois personnes interviewées, l'assistant vocal apporter une certaine sécurité supplémentaire et dispose également d'un nombre important de sources qui peuvent offrir une valeur ajoutée à des fins d'aider les personnes âgées et vulnérables chez elles.

4.3.3.2 Risques

Bien que l'assistant vocal procure bon nombre d'avantages, les risques liés à son utilisation dans des situations d'urgence à domicile se montrent conséquents également. En effet, les avancées actuelles présentent encore des lacunes amenant des risques de mauvais fonctionnement de la technologie.

Ahmed Bouzid pointe le risque que la transaction ne se réalise pas. Si la connexion internet ne fonctionne plus, que l'assistant vocal ne se déclenche pas, qu'il a une réaction trop lente, ou encore qu'il n'appelle pas les secours quand il y a la nécessité, le problème est important. Suivant cette dernière situation, Axel Legay affirme qu'il se peut que l'assistant vocal identifie mal une urgence. Le dispositif peut par exemple bien identifier que la personne est tombée, mais estime qu'il ne nécessite pas l'envoi des secours, mais qu'en réalité la personne s'est ouverte le crâne et il s'avère que la situation est plus grave que ne le pense l'assistant vocal. Monsieur Bouzid vient nuancer cet argument en signalant qu'il y a également le risque du faux positif qui peut entrer en jeu. Il est possible que l'assistant vocal appelle les secours en pensant

avoir identifié une situation d'urgence alors que rien de grave ne s'est produit. Ce même expert précise que c'est coûteux et que cela demande un effort inutile. Il identifie que, lié à cela, vient le risque d'ambiguïté. Le principe du bouton destiné à appeler les services de secours sur le collier exprime une requête claire. Il y a plus d'ambiguïté par la voix car l'assistant vocal n'est pas toujours en mesure d'identifier si la situation est réellement urgente. Le design de l'assistant vocal a alors toute son importance selon monsieur Bouzid, car les conséquences de l'échec sont importantes lorsque celui-ci est mal conçu, en particulier dans ce genre de circonstances délicates.

Axel Legay ajoute qu'il y a également le risque qu'un assistant vocal peut se faire hacker, en particulier s'il s'agit d'enceintes connectées de grosses entreprises comme Amazon. De plus, une fois que l'utilisateur investit dans ce genre de technologie vocale, il y a, selon lui, le risque qu'elle ne soit pas adoptée, ayant alors engendré des frais inutilement. Enfin, Ahmed Bouzid avance la perspective que l'appareil vocal est commun à la maison, signifiant qu'il ne tient pas compte de qui s'adresse à lui puisqu'il est communautaire. L'assistant vocal risque alors de ne pas savoir qui écouter, ce qui entraîne un autre problème d'ambiguïté et de compréhension.

Le développement non suffisant des assistants vocaux à l'heure actuelle constitue un risque considérable dans les situations d'urgence à domicile, car les conséquences peuvent être potentiellement dramatiques.

4.3.3.3 Adoption de l'assistant vocal (Demande)

Afin d'évaluer le potentiel de l'assistant vocal dans les services de secours à distance, il convient de se pencher sur les facteurs qui motivent et qui restreignent l'adoption de cette technologie par les personnes âgées ou souffrant d'un handicap vivant seules, ainsi que par leur famille. La question de si la technologie vocale doit être développée par de grandes entreprises ou par des startups spécifiques se pose également.

4.3.3.3.1 Facteurs clés d'adoption

Des critères doivent être atteints afin que l'utilisateur adopte réellement cette technologie vocale dans le cadre d'une urgence. Les trois personnes s'accordent sur le fait qu'il doit y avoir

un effort pédagogique pour pouvoir adopter l'assistant vocal. En effet, Ahmed Bouzid affirme que les utilisateurs sont tenus de savoir comment utiliser l'assistant vocal et ne pas simplement l'acheter et essayer de se débrouiller avec ultérieurement. Axel Legay est persuadé qu'il y a une bonne opportunité dans la présence d'appareils vocaux à domicile, mais qui ne sera adoptée que si la société fait l'effort pédagogique qui n'est pas suffisamment déployé pour l'instant. Il est indispensable d'éduquer la population plus âgée afin de les familiariser davantage avec la technologie.

Stanley Goffinet marque son accord avec cette idée en évoquant lui aussi que la nécessité d'éduquer les personnes âgées à la technologie est important afin qu'elles comprennent comment utiliser et s'adresser à un assistant vocal. Il pense également à une autre alternative qui est celle de rendre le système tellement simple qu'il n'y aurait plus besoin d'éducation car l'utilisation serait innée. Monsieur Bouzid soutient cette possibilité en estimant qu'il faut des efforts des deux côtés. Certes l'utilisateur doit avoir un minimum de capacité d'utilisation, mais l'application doit aussi être bien conçue pour réellement apporter une plus-value et être adoptée, car il s'agit d'une situation délicate, d'urgence.

En outre, Axel Legay et Stanley Goffinet partagent l'opinion commune que l'assistant vocal nécessite une personnalisation à l'utilisateur. Axel Legay considère que l'appareil doit être personnalisé, en termes de personnes de contact ainsi que de façon d'interagir. Selon lui, il convient d'avoir un assistant vocal pouvant interagir avec plusieurs types de personnes dites faibles, comme les personnes âgées ayant des difficultés à s'exprimer ou entendre, des personnes à mobilité réduite, ou encore des personnes sourdes et/ ou muettes par exemple. Il faut alors former l'assistant vocal et l'adapter à la personne en permanence. Monsieur Legay estime que c'est une condition nécessaire et que c'est indispensable pour rassurer la personne qui va utiliser l'assistant. Pour ce faire, Stanley Goffinet déclare qu'il est essentiel d'avoir un gestionnaire qui paramètre l'assistant vocal à la personne. Monsieur Legay explique cette nécessité par le fait que si personne ne peut aider le nouvel utilisateur à configurer son dispositif vocal, la probabilité qu'il l'adopte diminue. Cet expert exprime qu'il faut alors investir dans la personne qui va aider à la configuration, qui sera personnalisée.

D'autres facteurs peuvent encore entrer en jeu selon les trois personnes interrogées. Axel Legay évoque le fait qu'il est nécessaire d'avoir une transparence totale sur la façon dont les données

privées sont gérées. Le cas échéant, les utilisateurs n'auront pas confiance et seront plus réticents à disposer d'un assistant vocal au sein de leur maison. Ensuite, selon Ahmed Bouzid, il faut évaluer dans quelle mesure l'arrivée d'un assistant vocal dans la chaîne des services de secours va affecter les parties prenantes, à savoir les travailleurs au sein des services de santé, les utilisateurs de cette technologie et les proches des utilisateurs. Si ces derniers sont impactés positivement, l'adoption sera plus garantie. Stanley Goffinet le rejoint en quelque sorte sur cette idée en avançant que l'assistant vocal ne doit pas remplacer les anciens systèmes mais les compléter. Il estime en effet que l'appareil vocal doit constituer un outil supplémentaire destiné à aider les personnes en situation d'urgence de leur domicile.

Il est évident que bon nombre d'éléments sont à prendre en compte pour favoriser l'adoption des assistants vocaux au domicile des personnes âgées et vulnérables. Néanmoins, les barrières venant à l'encontre de cette adoption ne sont pas négligeables.

4.3.3.2 Facteurs limitant l'adoption

La présence d'assistants vocaux dans les situations d'urgence suscite de nombreuses réticences. La barrière technologique est soulevée par nos trois intervenants. Axel Legay explique que beaucoup de personnes craignent de ne pas être capables d'utiliser l'assistant vocal ou de ne pas savoir le configurer. Ceci va de pair avec la barrière pédagogique car, selon monsieur Legay, malgré les opportunités que présente l'assistant vocal, il subsiste un manque évident de pédagogie et d'explications. La nécessité d'éduquer les personnes âgées et de faire configurer l'appareil au préalable est d'autant plus importante. Ahmed Bouzid ajoute que, malgré la crainte technologique, la voix demeure tout de même, pour les personnes âgées et/ ou en difficulté, un moyen plus facile à utiliser que le tactile. De plus, Stanley Goffinet poursuit en précisant que, même si le dispositif est configuré, il est parfois compliqué de l'installer dans des foyers qui n'ont pas de connexion internet ou de téléphone portable, ce qui devient très coûteux lorsqu'il faut installer tout depuis zéro.

Ahmed Bouzid aborde également le problème de *endpointing*, qui revient pour chaque cas d'usage. Il explique que celui-ci peut s'appliquer en début et en fin de requête. En ce qui concerne le début, l'assistant vocal peut avoir cru entendre le mot déclencheur mais en réalité l'utilisateur a dit un terme proche de ce mot, déclenchant alors inutilement l'assistant vocal.

Concernant la fin, l'assistant vocal peut croire que la requête est finie lorsque l'utilisateur doit encore finir sa requête. Cela suscite une incompréhension et des malentendus, en particulier chez les personnes âgées, car elles ont tendance à parler de la même manière qu'à un humain, c'est-à-dire en prenant leur temps, en réfléchissant et en faisant des pauses. L'assistant vocal ne permet pas encore de reconnaître ces phénomènes et ne permet pas non plus de parler au-delà d'une certaine limite de temps. Les personnes non habituées à la technologie mettront plus de temps à comprendre cet inconvénient et seront plus susceptibles de rejeter l'assistant vocal.

Enfin, l'intrusion à la vie personnelle est l'un des obstacles à l'adoption mentionné par Axel Legay. Celui-ci explique que les utilisateurs ont donné un nombre important d'informations personnelles en configurant l'appareil vocal, ne serait-ce que les personnes de contact ainsi que leur manière d'interagir. L'expert ajoute que certaines personnes sont persuadées que leurs données sont diffusées et que des firmes qui possèdent ces données peuvent leur faire perdre des droits, augmentant leur assurance par exemple. Cette gestion des données personnelles est un frein pour beaucoup.

Les trois personnes que nous avons pu interroger évoquent des obstacles similaires et apportent chacune des informations complémentaires. Les freins restent importants dans le contexte des secours à distance.

4.3.3.3 Adoption de technologie développée par les grands acteurs ou des startups

Les avis relatifs à la technologie qui devrait être présente dans l'assistant vocal au sein des domiciles ne divergent guère. Les trois experts s'accordent pour dire que les technologies développées par des startups semblent les plus appropriées pour les situations d'urgence à domicile.

Ahmed Bouzid affirme que les assistants vocaux devraient être développés par des entreprises spécifiques à cet usage. Il estime que celles-ci iront plus en profondeur dans la spécialisation du cas des services de secours à distance, car elles font de la recherche exclusivement à cet effet, mais aussi car elles comprennent le problème dans son intégralité et s'engagent sur le long terme. Ainsi, monsieur Bouzid est convaincu qu'avoir recours à ce genre de technologie

apportera une meilleure valeur ajoutée et une meilleure expérience. Axel Legay partage également cet avis en émettant l'idée qu'il est plus opportun de rester sur quelque chose de restreint avec une technologie plus souveraine dans une telle situation délicate. Cet expert imagine des assistants vocaux développés par des petites startups belges dont les *clouds* sont maîtrisés, et en ayant les employés qui se rendent aux domiciles et qui assurent une transparence maximale. Il pense en effet que ces startups doivent partager leur façon de procéder ainsi que la manière dont les données sont gérées et stockées en concordance avec le RGPD. Monsieur Legay affirme que le recours à ce genre de technologie sera plus cher mais aussi plus sûr, point particulièrement important pour cette situation sensible.

Ces deux experts considèrent que les technologies des grands acteurs tels qu'Amazon ou Google sont moins appropriées pour les urgences à domiciles. Axel Legay explique que les assistants vocaux de ces géants technologiques fonctionnent avec des conditions d'amélioration, c'est-à-dire qu'ils enregistrent en permanence afin d'améliorer la reconnaissance vocale. Par ailleurs, Ahmed Bouzid souligne qu'Amazon, par exemple, ne cherche pas à se spécialiser mais à toucher le plus grand nombre de personnes possible. Il admet que c'est logique car c'est leur *business model*, mais que ce n'est pas l'idée recherchée pour les services d'urgence à domicile.

Quant à Stanley Goffinet, ce dernier estime que les deux options sont envisageables, mais que cela va dépendre des utilisateurs. En effet, il rejoint l'opinion des deux experts précédents en affirmant que la plupart des personnes demandent que la technologie soit locale afin de garder le contrôle sur leurs données. Cependant, si la personne n'a aucune crainte quant à la divulgation de ses données, monsieur Goffinet juge qu'elle peut recourir à n'importe quel système.

Majoritairement, il en ressort que la technologie vocale développée par des startups est plus adéquate pour le cas d'usage des services de secours à distance.

4.3.3.4 Prédiction futures

Les trois intervenants communiquent leur vision future quant à l'utilisation des assistants vocaux dans les situations d'urgence à distance, après avoir considéré tous les facteurs nécessaires.

Ahmed Bouzid semble être confiant sur le fait que les assistants vocaux vont être améliorés à des fins positives pour le contexte des secours à domicile. Il pense qu'ils ne seront pas seulement modifiés pour offrir plus de fonctionnalités dans un nombre croissant de situations, mais aussi pour mieux interagir avec les utilisateurs, où les soucis de *endpointing* et de reconnaissance vocale seront abordés en priorité. Grâce à ces progrès, Stanley Goffinet affirme que les personnes âgées seront plus enclines à les intégrer dans leur domicile et par conséquent à s'adapter et installer le nécessaire requis pour la mise en place d'un assistant vocal dans le foyer. Ce dernier est également convaincu qu'avec ces avancées, l'assistant vocal pourra traiter davantage de cas de figure qui pourront être utiles dans les situations d'urgence.

Ensuite, Axel Legay et Ahmed Bouzid conviennent tous deux que les assistants vocaux vont devenir de plus en plus proactifs. Le premier imagine l'assistant vocal transmettre des renseignements aux services de secours pour évaluer une situation. Par exemple, l'appareil pourra détecter lorsqu'il y a des anomalies dans le tableau médical de la personne, tout comme l'augmentation de fièvre, et pourra en avertir les services afin qu'ils puissent agir en conséquence. Monsieur Legay explique que le dispositif vocal pourrait alors se substituer à la personne de façon concrète et donner un certain nombre d'indications pour décider de la suite. Ahmed Bouzid estime quant à lui que les assistants vocaux arriveront à être plus proactifs lorsqu'ils maîtriseront la reconnaissance d'émotions. Il signale qu'il est déjà possible pour un assistant vocal de détecter certaines émotions mais sans qu'il soit en mesure de les exploiter. Stanley Goffinet partage cette vision de la possibilité de reconnaissance des émotions et affirme que cela se réalisera sur base du ton de la voix. Il ajoute que l'assistant vocal pourrait également capter le sentiment de la personne et lui-même refléter ce sentiment ainsi que ses émotions pour que la personne réponde plus facilement. Cela garantirait alors un dialogue plus fluide avec la personne physique.

Par ailleurs, Stanley Goffinet suppose que l'assistant vocal pourrait se trouver sur davantage de supports. En plus d'être présent sur le smartphone ou l'enceinte connectée de la personne, il pense qu'il pourrait notamment se trouver sur le frigo ou dans les systèmes auditifs de cette personne âgée par exemple. Selon monsieur Goffinet, le fait que l'appareil fasse partie d'un système domotique représente des opportunités car l'assistant vocal serait disponible sur des terminaux plus communs à la personne. Axel Legay envisage par ailleurs que les assistants vocaux pourraient être reliés avec des puces dans les corps des personnes âgées pouvant mesurer certains types d'information, par exemple pour les personnes diabétiques. Il doute cependant de la volonté de la population à adopter ce genre de dispositifs. Enfin, cet expert estime que les assistants vocaux ne pourront être présents dans les services d'urgence que de manière limitée, car il s'agit d'un secteur sensible. Il est cependant persuadé que cette technologie peut créer plus d'emplois qu'elle n'en supprime, à condition que l'ultime objectif soit de maximiser le confort et le bien-être des êtres humains, et non le profit.

Globalement, nos trois différents experts ont une imagination débordante d'usages que les assistants vocaux peuvent réaliser pour venir en aide aux personnes âgées et vulnérables. Cette technologie doit néanmoins être utilisée avec modération dans ce domaine particulièrement délicat.

4.3.4 Conclusion du cas d'usage

Compte tenu des données rassemblées lors des interviews réalisées avec ces trois personnes familières avec le cas d'usage des services d'interventions d'urgence à domicile, nous pouvons nous attendre à des progrès et des avancées significatives quant aux assistants vocaux dans ce milieu.

L'hypothèse préalablement réalisée pour ce cas d'usage défend la possibilité de l'assistant vocal de renforcer la sécurité des personnes en possession d'une enceinte connectée et de rendre plus efficaces les interventions en cas d'urgence à domicile. Après avoir analysé les opinions et perceptions des trois personnes interrogées, nous convenons de confirmer cette hypothèse. Tous les intervenants affirment que la présence d'un assistant vocal augmente la sécurité des personnes âgées comparativement à la situation où il n'y en a pas, et vient par conséquent rassurer les proches. Il est vrai que ce ne sera pas l'élément qui représente la source de sécurité

des personnes dites plus faibles, mais ajoute toutefois une sûreté en plus. En outre, d'après les commentaires de chaque intervenant, l'assistant vocal peut transmettre des informations importantes afin de rendre les interventions des services de secours plus efficaces. Les perspectives d'avenir élaborées par les personnes interrogées s'avèrent optimistes et préconisent le fait que les dispositifs vocaux seront de plus en plus proactifs afin d'assurer un service fluide et performant.

Néanmoins, il est évident que dans des situations d'urgence, l'intelligence artificielle pourra compléter et renforcer la présence humaine mais ne viendra jamais la remplacer. De plus, certains risques persisteront et certains aspects viendront toujours limiter l'adoption de la technologie vocale au sein des foyers, mais n'empêcheront pas que l'assistant vocal s'immiscera de manière toujours plus conséquente dans les services d'urgence à domicile.

CONCLUSION

Notre mémoire avait pour objectif de démontrer si les assistants vocaux apportent une valeur ajoutée dans nos trois cas d'usage. Notre question de recherche, « *L'état d'adoption de l'assistant vocal en fonction de sa valeur ajoutée apportée à trois cas d'usage majeurs : Aide à la conduite, apprentissage d'une langue étrangère, services de secours à distance* », peut dorénavant être répondue.

Trois hypothèses respectives à chaque cas avaient été élaborées. Deux d'entre elles peuvent être confirmées, tandis que la première a dû être réfutée. En effet, l'assistant vocal au sein des voitures ne permet pas encore d'apporter une réelle sécurité supplémentaire à la conduite.

Dans le contexte de la conduite, il est à constater que parmi les facteurs clés d'adoption, ce qui motive le consommateur à utiliser l'assistant vocal est davantage l'expérience retirée plutôt que l'aspect sécuritaire. En effet, les commandes vocales qui sont existantes, à ce jour, ne sont pas celles qui sont plus à même d'augmenter la sécurité à bord de la voiture. En revanche, celles qui sont amenées à voir le jour sont plus prometteuses en termes de sécurité. Une des raisons est que, jusqu'à présent, les assistants vocaux les plus connus du grand public sont ceux développés par des multinationales telles que Google ou Amazon. Une fois que les assistants vocaux embarqués, développés par les constructeurs automobiles, bénéficieront de la notoriété dont jouissent les GAFA, il est plus probable que l'aspect sécuritaire lié à la conduite ressorte.

De plus, la charge cognitive que prend l'assistant vocal dans le véhicule peut diminuer l'attention du conducteur, et par conséquent réduire son niveau de sécurité.

En ce qui concerne le côté technique, la technologie embarquée et la technologie externe présentent toutes deux des avantages, le plus approprié semble être un système hybride entre ces deux options même si la reconnaissance vocale de la technologie externe est susceptible d'être la plus performante. Toutefois, le choix réside dans le chef du constructeur automobile et peut dépendre de plusieurs facteurs, tels que ses moyens financiers, la segmentation client ou encore l'image de marque.

En raison des éléments cités ci-dessus, l'hypothèse de recherche du cas d'usage n°1, à savoir : « L'assistant vocal apporte une plus-value en termes de sécurité et d'expérience aux conducteurs de voitures qui en sont équipées » a été rejetée.

Dans le cadre de l'apprentissage d'une langue étrangère, l'assistant vocal présente beaucoup de bienfaits, en particulier dans les compétences de l'audition et de l'expression orale. Le fait de fournir un feedback est très utile pour pouvoir progresser lors de l'apprentissage d'une langue étrangère, ce que peut offrir un assistant vocal. Au contraire du professeur, qui n'est pas en mesure de donner un suivi particulier lorsqu'il enseigne à des groupes d'étudiants. De plus, l'assistant vocal assure une exposition à la langue beaucoup plus importante. Par ailleurs, au plus une personne est exposée à l'écoute de la langue, au plus elle se familiarise avec celle-ci. Si cela a déjà été démontré pour le visionnage de films, il en va de même pour l'exposition à l'assistant vocal. Cependant, force est de constater qu'il existe une fracture digitale au sein du corps enseignant. En effet, les utilisateurs qui souhaitent exploiter ces technologies ne les comprennent pas forcément et ne sont pas conscients de leurs limites. Cela constitue d'une part un réel problème dans la mise en œuvre de l'IA mais également un frein à l'adoption. Le manque de compréhension de certains accents par l'assistant vocal est également une limite à son utilisation. Globalement, ce dispositif assurera un apprentissage plus rapide, plus qualitatif et une autonomie pour l'étudiant. Il est cependant nécessaire que cet outil pédagogique soit utilisé tout en étant encadré par un professeur. De ce fait, l'assistant vocal sera un outil pédagogique sous réserve d'être utilisé en complémentarité avec l'enseignant.

Finalement, au sujet des services de secours à distance, l'assistant vocal peut fournir une sécurité supplémentaire, pour autant qu'un humain reste engagé dans le processus. La présence de l'assistant vocal complète le rôle des services de secours en fournissant et évaluant des données utiles relatives à l'utilisateur. Les interventions peuvent alors être plus efficaces et la présence d'un assistant vocal a tendance à rassurer les proches également. Même si l'assistant vocal ne va pas constituer la sûreté absolue de l'utilisateur, sa sécurité va toutefois être renforcée avec la présence d'un dispositif vocal. Outre la sécurité apportée, l'assistant vocal leur apporte une autonomie accrue, qui rend leur maison plus adaptée à leur vieillissement. Cependant, étant donné la situation délicate des interventions en cas d'urgence, les conséquences d'une éventuelle défaillance de l'assistant vocal peuvent être sérieuses, ce qui constitue un risque important à son utilisation. C'est pourquoi l'effort dans la conception, la simplification et la

personnalisation d'un assistant vocal ainsi que l'effort pédagogique des utilisateurs sont des éléments clés qui viendront favoriser l'adoption.

En raison des éléments cités ci-dessus, les hypothèses de recherche des cas d'usage n°2 et 3, qui sont respectivement : « L'usage de l'assistant vocal dans l'apprentissage d'une langue étrangère apporte une plus-value en tant qu'outil pédagogique » et « L'assistant vocal adopté renforcera la sécurité des personnes en possession d'une enceinte connectée et rendra plus efficaces les interventions en cas d'urgence à domicile » ont été acceptées.

Dans chaque cas d'usage, la question de la protection des données personnelles s'est également posée. Il s'avère que, quel que soit le cas d'usage, cela peut constituer un véritable frein. Les utilisateurs en sont conscients et la potentielle solution serait de se tourner vers des plus petites entreprises qui assureraient une transparence maximale à cette problématique. Un autre élément qui peut freiner les utilisateurs, en particulier les adultes, est la communication avec une machine, qui ne leur semble pas naturelle. Cela peut d'ailleurs créer un cercle vicieux : il a été constaté que lorsque la personne ne se fait pas comprendre par l'assistant vocal, elle a comme réflexe de changer son mode de communication, qui devient alors inapproprié, et par conséquent l'assistant vocal ne comprend pas. Cet ensemble de causes à effets dégrade l'utilisation de l'assistant vocal.

Enfin, en conclusion du sujet choisi pour notre mémoire, il apparaît que l'assistant vocal va présenter énormément d'opportunités futures dans chaque cas d'usage. Le niveau d'adoption de l'assistant vocal ne fera qu'accroître dans les prochaines années et la voix deviendra rapidement un standard.

Limites à la recherche

Il est intéressant de clôturer ce mémoire en exprimant notre avis sur l'objectivité et la pertinence du travail effectué. Dans le courant de ce mémoire, nous avons rencontré un certain nombre de limites relatives à notre méthodologie ainsi qu'aux choix réalisés pour notre partie empirique.

La première limite est liée à la méthodologie et plus précisément au choix de notre étude. De par son caractère qualitatif, il est impossible d'obtenir de résultats généralisés de notre étude (Pleyers, 2020). Dès lors, nos résultats ne sont pas quantifiables et ne permettent pas de réaliser une analyse statistique. Les interviews avec les experts constituent notre base de données mais il se peut que des erreurs d'interprétation surviennent. De plus, les guides d'entretien créés à cet effet ont été utilisés en guise de support. De ce fait, nous n'avons pas adressé les mêmes questions à tous les répondants.

La seconde limite encourue est apparue lorsque nous cherchions des experts à interviewer pour compléter notre échantillon. Il a été compliqué de trouver des personnes qui étaient expertes à la fois dans le domaine des cas d'usage et en matière d'assistant vocal. Nous avons d'ailleurs été contraintes d'annuler certaines interviews car nous nous rendions compte que la personne sollicitée n'était experte que dans un des deux aspects. Cela ne nous aurait pas été utile dans notre analyse. D'un autre côté, nous avons déploré un nombre important de refus ou de non-retours de la part de personnes contactées. En effet, le taux de personnes non volontaires s'élève à 88,88%.

Une troisième limite concerne notre deuxième cas d'usage, à savoir l'apprentissage d'une langue étrangère. Il se pourrait en effet qu'un biais existe au vu de celui-ci. Nous avons eu l'occasion d'interviewer trois personnes pour ce cas d'usage. Cependant, deux des personnes sont expertes sur le sujet de l'application développée par la startup belge, Flowchase. Leurs interviews étaient donc complémentaires car Robin Guérit est le développeur de l'application et François-Xavier Fievez est utilisateur. Tandis que la troisième personne, Maria Perillo, détenait son expertise dans l'utilisation d'Alexa. La comparaison des propos de ces intervenants n'est donc pas la plus objective et approfondie, ce qui montre que les profils des répondants ne sont pas très diversifiés de ce point de vue. Par ailleurs, certaines personnes ont un lien direct ou bien avec les GAFAs ou des startups. Dans les deux cas, il n'est pas forcément dans leur

intérêt d'émettre certains faits et les résultats obtenus pourraient varier considérablement en fonction de la position des personnes interviewées. Il aurait alors été intéressant d'interroger une autre personne qui utilise ou qui travaille pour l'un des grands acteurs tels que Amazon ou Google, pour pouvoir assurer une comparaison plus exhaustive.

Une autre limite rencontrée est le fait que l'on s'est rendues compte, après avoir fait l'analyse de nos cas, que nous avons été au-delà de nos hypothèses. Au fil de la discussion, les experts ont amené des problématiques plus larges au sujet des assistants vocaux. Bien que très intéressantes, ces discussions nous ont quelque peu éloigné de la question principale de la plus-value. L'hypothèse ne reprend en effet qu'une partie de l'analyse.

Enfin, le manque de données scientifiques sur le sujet étudié a été une limite à la réalisation de notre revue de littérature. En effet, alors que le sujet des assistants vocaux a connu un essor récent, le nombre d'études antérieures pertinentes sur le sujet est limité. Néanmoins, comme il s'agit d'un mémoire analysant trois cas d'usage différents, celui-ci fournit une vue hélicoptère des cas.

Finalement, nous aurions souhaité réaliser une analyse par scénarios. Une telle analyse aurait permis de schématiser les potentielles prédictions futures pour le marché des assistants vocaux. Cependant, le temps était contre nous. Nous avons donc dû clôturer ce mémoire sur notre analyse, laissant l'analyse des scénarios.

Suggestions pour recherches futures

L'intérêt de cette recherche était d'évaluer la valeur que peut ajouter l'assistant vocal selon le niveau d'adoption de la part des utilisateurs dans trois différents cadres. Afin d'approfondir le sujet, plusieurs pistes futures pourraient être intéressantes à emprunter.

Pour des recherches futures plus approfondies, l'analyse en profondeur de l'un de nos trois cas d'usage et aller plus loin dans les conclusions pourrait constituer une piste à traiter. Se concentrer uniquement sur un cas d'usage et dès lors constituer un échantillon plus complet et diversifié peut faire office d'une recherche plus exhaustive. Cela déboucherait sur une étude plus complète et détaillée.

Il serait également intéressant de réitérer cette recherche dans quelques années. En effet, le manque d'articles scientifiques et l'adoption des assistants vocaux n'étant pas encore considérablement développée à l'heure actuelle constitue une limite. Néanmoins, l'évolution de ces technologies se faisant drastiquement et rapidement, examiner la façon dont leur adoption a évolué dans les trois différents cas d'usage peut constituer un sujet clé.

Finalement, rappelons que nous avons dû faire un choix parmi les cas d'usage suggérés lors de la formulation de notre question de recherche (voir chapitre 1). Bien que cela nous ait aidé à cloisonner nos recherches, nous avons dû abandonner l'idée d'étudier certains cas d'usage. Il serait alors intéressant de faire des recherches sur l'un ou plusieurs des cas d'usage que nous avons choisi de ne pas traiter. En effet, analyser la valeur que l'assistant vocal apporte dans les voitures autonomes dans l'accompagnement d'enfants, dans le domaine de la traduction, dans le secteur des soins de santé en général, ou encore dans la maison intelligente peut faire objet de piste très intéressante pour un futur mémoire.

Points de vue personnels

Notre vision personnelle du futur est que les assistants vocaux vont être adoptés par le grand public. Nous pensons que la voix est une interface plus simple que celles qui la précèdent, ce qui va favoriser l'adoption par une population considérablement plus large. Par conséquent, nous estimons que la présence des assistants vocaux dans nos trois cas d'usage distincts va s'amplifier davantage dans les prochaines années. Nous émettons également notre avis personnel relatif à l'évolution de chacun de nos cas d'usage.

Dans le cadre de la conduite, nous pensons que la reconnaissance des états émotionnels du conducteur va constituer une opportunité et sécurité significative. Le fait que l'assistant vocal va pouvoir détecter l'état de fatigue ou toute autre émotion sur base du ton de la voix pourra amener le conducteur à prendre conscience du danger et à prendre des mesures sécuritaires appropriées.

Ensuite, utilisé dans l'apprentissage d'une langue, l'assistant vocal devrait apporter son aide. Un certain temps d'adaptation sera encore nécessaire avant de voir cette technologie introduite

au sein des cours de langues. Il serait important de conscientiser les enseignants avant de lever certaines craintes. Nous considérons que l'assistant vocal devra être utilisé comme outil pédagogique complémentaire et non de substitution.

Finalement, dans le contexte de l'aide aux personnes à domicile, nous estimons qu'avec les évolutions technologiques, les appareils seront de plus en plus simplifiés à l'utilisation. Par ailleurs, comme l'évolution des interfaces nous l'indique (voir section 2.1.1), la complexité des technologies s'estompe, ce qui favorisera et accélèrera l'adoption pour les personnes non-habituées à la technologie.

De manière générale, nous pensons, comme une partie des interviewés, qu'il y a un lien de cause à effet entre le manque d'éducation technologique et l'adoption des assistants vocaux. Dans le corps enseignant, la fracture digitale est présente. Au niveau des seniors, cet écart se fait ressentir davantage. Cependant, nous pensons que de nombreux aménagements vont être de plus en plus mise en place afin que l'utilisation de la technologie soit accessible à tous. Par exemple, on pourrait envisager que l'éducation technologique soit donnée dans l'enseignement secondaire. Ainsi, les utilisateurs pourraient comprendre les limites de cette technologie. À partir d'un certain moment donné, les populations vieillissantes feront partie des *digital natives* et cela résoudra ce problème.

Ce travail de recherche et les interviews conduites dans le cadre de ce mémoire furent très intéressantes. Nous avons eu l'honneur de rencontrer des personnes expertes qui nous ont apporté leur avis et connaissance approfondie en la matière. L'analyse de ces trois cas nous a permis d'avoir une vue d'ensemble sur le potentiel du marché des assistants vocaux.

BIBLIOGRAPHIE

- Abela, A. (2019, 27 janvier). *Cliquer, toucher, parler : l'évolution de la navigation*. mydigitalweek. <https://mydigitalweek.com/cliquer-toucher-parler-levolution-de-la-navigation/>
- Académie de Caen. (s.d.). *Le système phonologique du Français*. https://www.ac-caen.fr/dsden50/circo/mortain/IMG/pdf/2._systeme_francais.pdf
- Activate. (2018). *Activate Tech & Media Outlook 2018*. November 2018, 153. <https://fr.slideshare.net/ActivateInc/activate-tech-media-outlook-2018>
- Aeris. (2020, 10 avril). *What is IoT ? Defining the Internet of Things (IoT)*. Aeris | India. <https://www.aeris.com/in/what-is-iot/>
- AI 4 Belgium. (2019). *AI 4 Belgium*. https://www.ai4belgium.be/wp-content/uploads/2019/04/report_en.pdf
- Aleksic, P., Michaely, A. H., Parada, C., Simko, G., & Zhang, X. (2017). Keyword spotting for Google assistant using contextual speech recognition. *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 272-278. <https://doi.org/10.1109/asru.2017.8268946>
- Alimi, J. (s. d.). *L'assistant vocal, l'interface homme-machine du futur*. Suricats Consulting. <https://www.suricats-consulting.com/expertise-commerce-assistant-vocal/>
- Al-Kaisi, A. N., Arkhangelskaya, A. L., & Rudenko-Morgun, O. I. (2020). The didactic potential of the voice assistant “Alice” for students of a foreign language at a university. *Education and Information Technologies*, 26(1), 715-732. <https://doi.org/10.1007/s10639-020-10277-2>
- Amberscript. (2021, 14 juin). *How Does Speech to Text Software Work*. Amberscript. <https://www.amberscript.com/en/blog/how-speech-to-text-software-works/>

- Aunege. (s.d.). *Marketing fondamental - Les différents types de demande*.
http://ressources.aunege.fr/nuxeo/site/esupversions/83e876d5-3c45-45cb-a888-2af03045ca8e/co/L1_3_1_differeents_types_demande.html
- Barrow, E. (2019, 11 avril). *How Voice Technology Can Support Education & Learning*. 3 SIDED CUBE. <https://3sidedcube.com/how-voice-technology-can-support-education-learning/>
- Beaudouin, M. (2016, 1 juillet). *50 ans d'interaction homme-machine : retours vers le futur*. Interstices. <https://interstices.info/50-ans-dinteraction-homme-machine-retours-vers-le-futur/>
- Ben-Yaacov, G., Stone, E., & Goldman, R. (2001). Applying Moore's technology adoption life cycle model to quality of EDA software. *Proceedings of the IEEE 2001. 2nd International Symposium on Quality Electronic Design*. Published. <https://doi.org/10.1109/isqed.2001.915209>
- Bergounhoux, J. (2018, 19 mars). *L'automobile, un marché clé pour les assistants vocaux*. L'usine digitale. <https://www.usine-digitale.fr/article/l-automobile-un-autre-marche-cle-pour-les-assistants-vocaux.N664964>
- Bertrand., J.-M. (1987). *Test de Turing : jeu d'imitation ou test d'intelligence ?* <https://doi.org/10.3406/quad.1987.2095>
- Bhasin, H. (2020, 28 février). *Different Types of Demand*. Marketing91. <https://www.marketing91.com/types-of-demand-2/>
- Blanchot, V. (2018, août 20). *Histoire de l'intelligence artificielle*. Siècle Digital. <https://siecldigital.fr/2018/08/20/histoire-intelligence-artificielle/>

- Black, G. (2019, 26 septembre). *Are you done? ? ? ? Endpoint Detection and Conversational Rhythm with Voice Assistants*. LinkedIn. <https://www.linkedin.com/pulse/you-done-endpoint-detection-conversational-rhythm-voice-grace-black/>
- Brard C. & Eraly H. (2017). *Les assistants vocaux, une aubaine pour les seniors ?* Énéo Focus, 2017/26.
https://www.eneo.be/images/analyses/2017/201726_assistants_vocaux_aubaine_pour_les_seniors.pdf
- Brush, K. (2019, 17 octobre). *Digital Ecosystem*. SearchCIO. <https://searchcio.techtarget.com/definition/digital-ecosystem>
- Caizergues, G. (2020, 23 juillet). *Les assistants vocaux ont du mal à comprendre les accents*. Enregistrersous.com. <https://www.enregistrersous.com/les-assistants-vocaux-ont-du-mal-a-comprendre-les-accents/>
- Calvary, G., & Coutaz, J. (2007). Métamorphose des IHM et Plasticité. *Revue d'Interaction Homme-Machine*, 1.
- Caring Home. (s.d.). *The Best Voice-Activated Medical Alert Devices*. Caring.com. <https://www.caring.com/best-medical-alert-systems/best-voice-activated-systems/>
- Chapuzet, A. (2021, 17 juin). *Assistant vocal RGPD : Peut-on assurer le respect des données ?* Vivoka. <https://vivoka.com/fr/assistant-vocal-rgpd/>
- Chapuzet, A. (2019, 2 décembre). *Reconnaissance vocale : Comment ça marche ?* LinkedIn. <https://www.linkedin.com/pulse/reconnaissance-vocale-comment-%25C3%25A7a-marche-avec-vivoka-aur%25C3%25A9lien-chapuzet/?trackingId=Bd2LzJkLSmq%2BKcUEJjHJSQ%3D%3D>
- Claude, G. (2019, 30 octobre). *Étude qualitative : définition, techniques, étapes et analyse* [Diapositives]. Scribbr. <https://www.scribbr.fr/methodologie/etude-qualitative/>

CNIL. (2020, septembre). *À votre écoute - Exploration des enjeux éthiques, techniques et juridiques des assistants vocaux.*
https://www.cnil.fr/sites/default/files/atoms/files/cnil_livre-blanc-assistants-vocaux.pdf

CNIL. (s.d.). CNIL. LinkedIn. <https://www.linkedin.com/company/cnil---commission-nationale-de-l%27informatique-et-des-libert-s>

CNRS. (2012, mai). *L'héritage d' Alan Turing.* <http://www.cnrs.fr/fr/pdf/jdc/Turing.pdf>

Cohen, D. (2021, 25 mars). *L'explosion de l'expérience vocale : format "Hype" du moment ou stratégie prometteuse pour les marques ?* Tell Me More.
<https://www.tellmemore.agency/blog/lexplosion-de-lexperience-vocale-format-hype-du-moment-ou-strategie-prometteuse-pour-les-marques/>

Commande vocale. (2019, 16 octobre). *Un assistant vocal comme outil pédagogique en classe d'école.* Commande vocale. <https://commandevocale.fr/assistant-vocal-outil-pedagogique-classe-d-ecole/>

Conseil de l'Europe. (s.d.). *L'IA, c'est quoi ?* <https://www.coe.int/fr/web/artificial-intelligence/what-is-ai>

CSA. (s. d.). *Le CSA.* CSA Belgique. <https://www.csa.be/le-csa/>

Damacharla, P., Devabhaktuni, V. K, Dhakal, P., Javaid, A. Y. & Vege, H. K. (2020). IVACS : Intelligent Voice Assistant for Coronavirus Disease (COVID-19) Self-Assessment. *2020 International Conference on Artificial Intelligence & Modern Assistive Technology (ICAIMAT)*. <https://doi.org/10.1109/icaimat51101.2020.9308013>

DataScientest. (2021, juillet 12). *Archives des Deep Learning.* <https://datascientest.com/category/deeplearning>

Dictionary.com. (2021, 19 janvier). *IBM.* <https://www.dictionary.com/e/acronyms/ibm/>

- Dilmegani, C. (2021, 5 juillet). *Natural Language Generation (NLG): What it is & How it works*. AIMultiple. <https://research.aimultiple.com/nlg/>
- Dosquet, E., Gouliaéva, O., & Moysan, Y. (2020). *La révolution des assistants vocaux - Comprendre les enjeux et réussir ses stratégies marketing : Comprendre les enjeux et réussir ses stratégies marketing (Marketing/Communication) (French Edition)*. DUNOD.
- Edu, J. S., Suarez-Tangil, G., & Such, J. M. (2020, août). *Smart Home Personal Assistants: A Security and Privacy Review*. <https://arxiv.org/pdf/1903.05593.pdf>
- edureka ! (2018, 29 mars). *AI vs Machine Learning vs Deep Learning | Machine Learning Training with Python | Edureka* [Vidéo]. YouTube. <https://www.youtube.com/watch?v=WSbgixdC9g8>
- EMnify. (2021, 22 avril). *What Is 4G ? Fourth Generation Cellular Networks Explained*. <https://www.emnify.com/en/resources/4g>
- Epstein, R., Roberts, G., & Beber, G. (2009). *Parsing the Turing Test*. Springer. <https://link.springer.com/content/pdf/10.1007%2F978-1-4020-6710-5.pdf>
- Frankenfield, J. (2021, 8 mars). *Artificial Intelligence (AI)*. Investopedia. <https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp>
- Frankenfield, J. (2020, 9 décembre). *AI Winter Definition*. Investopedia. <https://www.investopedia.com/terms/a/ai-winter.asp>
- Frankenfield, J. (2020, 11 novembre). *General Data Protection Regulation (GDPR)*. Investopedia. <https://www.investopedia.com/terms/g/general-data-protection-regulation-gdpr.asp>

Frankenfield, J. (2020, août 31). *Introduction to Natural Language Processing (NLP)*. Investopedia. <https://www.investopedia.com/terms/n/natural-language-processing-nlp.asp>

Frankenfield, J. (2020, 28 août). *Strong AI*. Investopedia. <https://www.investopedia.com/terms/s/strong-ai.asp>

Frankenfield, J. (2021, 25 février). *Weak AI*. Investopedia. <https://www.investopedia.com/terms/w/weak-ai.asp>

Future of Tech. (s.d). *The Technologies Behind Artificial Intelligence*. CompTIA's Future of Tech. <https://www.futureoftech.org/artificial-intelligence/3-technologies-behind-ai/#section-1>

Futura Tech. (s.d.). *Chatbot : qu'est-ce que c'est ?* Futura. <https://www.futura-sciences.com/tech/definitions/internet-chatbot-15778/>

Futura Tech. (s.d.). *Intelligence artificielle : qu'est-ce que c'est ?* Futura. <https://www.futura-sciences.com/tech/definitions/informatique-intelligence-artificielle-555/>

Futura Tech. (s.d.). *iOS : qu'est-ce que c'est ?* Futura. <https://www.futura-sciences.com/tech/definitions/multimedia-ios-15211/>

Futura Tech. (s. d.). *Système d'exploitation : qu'est-ce que c'est ?* Futura. <https://www.futura-sciences.com/tech/definitions/informatique-systeme-exploitation-11820/>

Gailly, B. (2020). *Identifying innovation opportunities*. Innovation management. (LLSMS2040)

Gallay, E. (2019, 28 octobre). *La domotique, c'est quoi ?* Idelec Plus. <https://www.idelecplus.com/blog/la-domotique-cest-quoi>

Geotab. (2021, 22 mai). *What is GPS?* <https://www.geotab.com/blog/what-is-gps/>

- Goffinet, S., & Schmitz, D. (2021). *Controlling security rules using natural dialogue : An application to smart home care*. <http://hdl.handle.net/2078.1/thesis:30519>
- Gong, J., Lu, J., Wu, L., & Zhang, T. (2016). Robot-assisted intelligent emergency system for individual elderly independent living. *2016 IEEE Global Humanitarian Technology Conference (GHTC)*, 628-633. <https://doi.org/10.1109/ghtc.2016.7857344>
- Guerraoui, R., & Hoang, N. L. (2020). *Turing à la plage - L'intelligence artificielle dans un transat : L'intelligence artificielle dans un transat (A la plage) (French Edition)*. DUNOD.
- Hadopi & CSA. (2019, mai). *ASSISTANTS VOCAUX ET ENCEINTES CONNECTÉES - l'impact de la voix sur l'offre et les usages culturels et médias*. https://www.hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2019_05_24_Assistants_vocaux_et_enceintes_connectees_FINAL.pdf
- Hargrave, M. (2021, 17 mai). *Deep Learning*. Investopedia. <https://www.investopedia.com/terms/d/deep-learning.asp>
- Heller, M. (2020, 6 janvier). *Deep learning vs. machine learning : Understand the differences*. InfoWorld. <https://www.infoworld.com/article/3512245/deep-learning-vs-machine-learning-understand-the-differences.html>
- Helmer, T. (2021, 9 avril). *In-car digital assistants hold a key to a delightful user experience*. Star. <https://star.global/posts/in-car-digital-assistants/>
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1), 81-88. <https://doi.org/10.1080/02763869.2018.1404391>

- Huber, A. (2019, 15 septembre). *Assistants vocaux : La Voice Tech pour l'autonomie des seniors ?* Medium. <https://medium.com/@albanhuber/assistants-vocaux-la-voice-tech-pour-lautonomie-des-seniors-484781d2a3d4>
- IA School. (2020, 16 avril). *Assistants vocaux : ces outils du quotidien (et du confinement) !* <https://www.intelligence-artificielle-school.com/actualite/assistants-vocaux-outils-quotidien-confinement/>
- IA School. (2021, août 3). *Intelligence Artificielle : quelles technologies nous cache-t-elle ?* https://www.intelligence-artificielle-school.com/actualite/intelligence-artificielle-quelles-technologies-nous-cache-t-elle%e2%80%89/?fbclid=IwAR0GKX0Y4B1vKT8rnbXYHj0nyEsqor7dxHoa_OiNzSZ8lqodb_OQmt1lqyk
- IONOS Digital Guide. (2019, 9 avril). *Qu'est-ce que le traitement automatique du langage naturel ?* <https://www.ionos.fr/digitalguide/web-marketing/vendre-sur-internet/le-traitement-automatique-du-langage-naturel-taln/>
- Ioualalen, A. (2021, 20 mai). *Les assistants vocaux sont-ils sur le point de devenir vos meilleurs interlocuteurs ?* LinkedIn. <https://www.linkedin.com/pulse/les-assistants-vocaux-sont-ils-sur-le-point-de-vos-arnault-ioualalen/?trackingId=XvozhrfjSXGpWqQc10f4xQ%3D%3D>
- Jacquemin, A. (2017). *Focus sur la méthodologie qualitative de recherche*. Séminaire d'accompagnement du mémoire. (LLSMD2094)
- Janev, V., Lakshen, G. A., & Vraneš, S. (2016, November). Big data and quality: A literature review. In *2016 24th telecommunications forum (TELFOR)* (pp. 1-4). IEEE. <https://doi.org/10.1109/telfor.2016.7818902>
- Jean, S. (2016, 21 novembre). *Apprendre une langue grâce aux intelligences artificielles*. Thot Cursus. <https://cursus.edu/articles/36478/apprendre-une-langue-grace-aux-intelligences-artificielles>

Kamath, U., Liu, J., & Whitaker, J. (2019). *Deep learning for NLP and speech recognition* (Vol. 84). Cham: Springer.

Kavlakoglu, E. (2020, 12 novembre). *NLP vs. NLU vs. NLG : the differences between three natural language processing concepts*. IBM - Watson. <https://www.ibm.com/blogs/watson/2020/11/nlp-vs-nlu-vs-nlg-the-differences-between-three-natural-language-processing-concepts/>

Kinsella, B. (2017, 5 novembre). *Gartner Hype Cycle Suggests Another AI Winter Could Be Near*. Voicebot.Ai. <https://voicebot.ai/2017/11/05/gartner-hype-cycle-suggests-another-ai-winter-near/>

Kinsella, B. (2020). *In-Car Voice Assistant Users Show Different Patterns Than on Smart Speakers with Making a Phone Call the Top Request*. Voicebot. <https://voicebot.ai/2020/02/24/in-car-voice-assistant-users-show-different-patterns-than-on-smart-speakers-with-making-a-phone-call-the-top-request/>

Kinsella, B., & Mutchler, A. (2020, janvier). *In-Car Voice Assistant Consumer Adoption Report*. Voicebot.ai. http://voicebot.ai/wp-content/uploads/2020/02/in_car_voice_assistant_consumer_adoption_report_2020_voicebot.pdf

Kinsella, B., & Mutchler, A. (2018, novembre). *Voice Assistant Consumer Adoption Report*. Voicebot. <https://voicebot.ai/wp-content/uploads/2018/11/voice-assistant-consumer-adoption-report-2018-voicebot.pdf>

Kiwak, K. (2020, septembre 25). *Speech Recognition*. SearchCustomerExperience. <https://searchcustomerexperience.techtarget.com/definition/speech-recognition>

Kopp, C. M. (2020, 3 juillet). *Understanding Business Models*. Investopedia. <https://www.investopedia.com/terms/b/businessmodel.asp>

Le Cun, Y. (2019). *Quand la machine apprend: la révolution des neurones artificiels et de l'apprentissage profond*. Odile Jacob.

LinkedIn. (s.d.). IBM. <https://www.linkedin.com/company/ibm/about/>

Lovens, P. (2020, 13 mars). *Flowchase, l'application qui va vous aider à devenir « fluent speaker »*. La Libre. <https://www.lalibre.be/economie/entreprises-startup/2020/03/13/flowchase-lapplication-qui-va-vous-aider-a-devenir-fluent-speaker-I3JT7ZVKHND5FOIXYM77CUVH3A/>

Marquenis, P., Papini, O., & Prade, H. (s.d.). *Turing et l'Intelligence Artificielle*. <http://audentia-gestion.fr/Intelligence-Artificielle/PDF/prade.pdf>

Marr, B. (s. d.). *What Really Is Natural Language Generation And Processing?* Bernard Marr & Co. <https://bernardmarr.com/default.asp?contentID=1293>

Martcorena-Sánchez, R., Ochoa-Orihuel, J., & Sáiz-Manzanares, M. (2020). Effectiveness of Using Voice Assistants in Learning: A Study at the Time of COVID-19. *International Journal of Environmental Research and Public Health*. Published. <https://doi.org/10.3390/ijerph17155618>

McCaffrey, M., Hayes, P., Hobbs, M., & Wagner, J. (2018). *Consumer Intelligence Series : Prepare for the voice revolution*. PwC. <https://www.pwc.com/us/en/advisory-services/publications/consumer-intelligence-series/voice-assistants.pdf>

Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughi, V. (2020). Advanced Methods for Photovoltaic Output Power Forecasting : A Review. *Applied Sciences*, 10(2), 487. <https://doi.org/10.3390/app10020487>

Mercedes-Benz. (s.d.). *How Do I Use MBUX ? | What is Mercedes-Benz User Experience?* Mercedes-Benz of Rockville Centre. <https://www.mbrvc.com/how-to-use-mbux/>

- Minker, W. & Néel, F. (2002). Développement des technologies vocales. *Le travail humain*, 65, 261-287. <https://doi.org/10.3917/th.653.0261>
- Minker, W., Schmidt, M., Stier, D., & Werner, S. (2019). *Exploration and Assessment of Proactive Use Cases for an In-Car Voice Assistant*. http://www.essv.de/pdf/2019_148_155.pdf
- Mutchler, A. (2018, 19 octobre). *VA-technology-adoption-curve-01*. Voicebot.Ai. <https://voicebot.ai/2018/10/19/phase-one-of-the-voice-assistant-era-is-over-long-live-phase-two/va-technology-adoption-curve-01/>
- Nguyen, T. (2021b, janvier 18). *4 Impactful Technologies From the Gartner Emerging Technologies and Trends Impact Radar for 2021*. Gartner. <https://www.gartner.com/smarterwithgartner/4-impactful-technologies-from-the-gartner-emerging-technologies-and-trends-impact-radar-for-2021/>
- Nicole, O. (2020, juin). *Voice Assistants and Privacy Issues*. Privacy Policies. https://www.privacypolicies.com/blog/voice-assistants-privacy-issues/#The_Implications_Of_The_Gdpr_For_Voice_Assistants
- Oh Yeah Sarah. (2019, 19 juillet). *Can you use voice assistants to practise speaking English?* Medium. <https://oh-yeah-sarah.medium.com/can-you-use-voice-assistants-to-practise-speaking-english-f2cb679941ee>
- Petropoulos, G. (2017, 6 avril). *Machines that learn to do, and do to learn: What is artificial intelligence?* Bruegel. <https://www.bruegel.org/2017/04/machines-that-learn-to-do-and-do-to-learn-what-is-artificial-intelligence/>
- Pleyers, G. (2020). *Marketing Research*. (LLSMS2000)
- Regan, T. (2017, août 11). *Home Health Care Makes Inroads With Amazon Alexa Pilot*. Home Health Care News. <https://homehealthcarenews.com/2017/08/home-health-care-makes-inroads-with-amazon-alexa-pilot/>

- Reynoso, R. (2019, mai 25). *A complete history of Artificial Intelligence*. G2. <https://www.g2.com/articles/history-of-artificial-intelligence>
- Roland Berger & Viseo. (2019, juillet). *La voix monte le son - La révolution naissante des assistants vocaux*. <https://www.rolandberger.com/fr/Insights/Publications/La-r%C3%A9volution-naissante-des-assistants-vocaux.html>
- Satratzemi, M., & Terzopoulos, G. (2020). Voice Assistants and Smart Speakers in Everyday Life and in Education. *Informatics in Education*, 473-490. <https://doi.org/10.15388/infedu.2020.21>
- Segal, T. (2021, 1 janvier). *Big Data*. Investopedia. <https://www.investopedia.com/terms/b/big-data.asp>
- SoundHound. (2020, 15 octobre). *How Drivers are Actually Using Voice Assistants in Cars*. SoundHound. <https://voices.soundhound.com/how-drivers-are-actually-using-voice-assistants-in-cars-bb1e3972301b/>
- SoundHound. (2020, 28 septembre). *How Voice Assistants are Shaping the Future of In-Vehicle Control*. SoundHound. <https://voices.soundhound.com/how-voice-assistants-are-shaping-the-future-of-in-vehicle-control/>
- Statista. (2021). *In-car voice assistant consumer use cases worldwide as of 2019*. <https://www.statista.com/statistics/1073545/worldwide-in-car-voice-assistant-use-cases-consumers/>
- Statista. (2021, 22 janvier). *Number of digital voice assistants in use worldwide 2019–2024*. <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/#statisticContainer>
- Steele, C. (2016, 4 novembre). *Personal Digital Assistant (PDA)*. SearchMobileComputing. <https://searchmobilecomputing.techtarget.com/definition/personal-digital-assistant>

- Techopedia. (2016, 26 juillet). *Backup*. Techopedia.Com.
<https://www.techopedia.com/definition/1056/backup>
- Techopedia. (2014, 18 juillet). *CarPlay*. Techopedia.Com.
<https://www.techopedia.com/definition/30204/carplay>
- Techopedia. (2012, 7 juin). *Text to Speech (TTS)*. Techopedia.Com.
<https://www.techopedia.com/definition/23843/text-to-speech-tts>
- TechTarget Contributor. (2021, 16 juin). *Natural Language Understanding (NLU)*. SearchEnterpriseAI. <https://searchenterpriseai.techtarget.com/definition/natural-language-understanding-NLU>
- Thurier, R. (2020, août 11). *Innovation et diffusion du call tracking auprès des entreprises en France*. Dexem. <https://blog.dexem.com/diffusion-innovation-call-tracking-entreprises-france/>
- Tomaz, C. (2019, 17 septembre). *Hype Cycle Gartner 2019 : quelle techno est au sommet de la hype cette année ?* HUB Institute - Digital Think Tank. <https://hubinstitute.com/2019/data-tendances-hypecycle-gartner-technologies-emergentes-ia-ecosysteme-capteurs-analytics>
- Touma, R. (2021, 24 février). *Clubhouse app: what is it and how do you get an invite to the exclusive audio app?* The Guardian. <https://www.theguardian.com/technology/2021/feb/17/clubhouse-app-invite-what-is-it-how-to-get-audio-chat-elon-musk>
- UKO. (s. d.). *Système embarqué : Définition*. Consulté le 6 août 2021, à l'adresse <https://www.ukonline.be/cours/embeddedsystems/programming/chapitre1-1>
- Verani, E. (2020, 24 avril). *IA symbolique vs Machine Learning appliqué au TALN*. Inbenta. <https://www.inbenta.com/fr/blog/ia-symbolique-vs-machine-learning-taln/>

Verchère, A. (2018, 2 mai). *Quels sont les usages des assistants vocaux ?* Siècle Digital.
<https://siecledigital.fr/2018/05/02/quels-sont-les-usages-des-assistants-vocaux/>

Walch, K. (2018, 30 novembre). *Intelligence Artificielle : quelle différence entre NLP et NLU ?* LeMagIT. <https://www.lemagit.fr/conseil/Intelligence-Artificielle-quelle-difference-entre-NLP-et-NLU>

Winston, P. (2010). *Lecture 14: Learning: Sparse Spaces, Phonology*. Lecture Videos | Artificial Intelligence | Electrical Engineering and Computer Science | MIT OpenCourseWare. <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/lecture-videos/lecture-14-learning-sparse-spaces-phonology/>

Wikipedia contributors. (2021, août 5). *Android Auto*. Wikipedia.
https://en.wikipedia.org/wiki/Android_Auto

Wikipedia contributors. (2021, 30 mai). *Diplôme d'études approfondies*. Wikipedia.
https://fr.wikipedia.org/wiki/Dipl%C3%B4me_d%27%C3%A9tudes_appfondies

Wikipedia contributors. (2021, 14 mars). *Enceinte connectée*. Wikipedia.
https://fr.wikipedia.org/wiki/Enceinte_connect%C3%A9e

Wikipedia contributors. (2021, août 2). *Massachusetts Institute of Technology*. Wikipedia.
https://en.wikipedia.org/wiki/Massachusetts_Institute_of_Technology

Wikipedia contributors. (2020, 11 novembre). *Miniaturisation*. Wikipedia.
<https://fr.wikipedia.org/wiki/Miniaturisation>

Wikipedia contributors. (2021, février 14). *Phonographe*. Wikipedia.
<https://fr.wikipedia.org/wiki/Phonographe>

Zhang, T., Shao, Y., Wu, Y., Geng, Y., & Fan, L. (2020). An overview of speech endpoint detection algorithms. *Applied Acoustics*, 160, 107133. <https://doi.org/10.1016/j.apacoust.2019.107133>

ANNEXES

ANNEXE 1: Glossaire

Mot	Définition
1. 4G	« La 4G est la quatrième génération de réseaux cellulaires. Rendus publics pour la première fois en 2009, les réseaux 4G offrent des débits de données nettement plus rapides, une latence plus faible et une utilisation plus efficace du spectre des fréquences radioélectriques. Aujourd'hui, plus de 50 % de toutes les connexions mobiles sont effectuées sur des réseaux 4G. » (EMnify, 2021).
2. Agent conversationnel	« Un chatbot, aussi appelé « agent conversationnel », est un programme informatique capable de simuler une conversation avec un ou plusieurs humains par échange vocal ou textuel. » (Futura Tech, s.d.).
3. Algorithme	« Suite d'instructions à exécuter. Le plus souvent, ces instructions sont exécutables par un ordinateur et incluent des opérations mathématiques, des tests, des boucles, etc. » (Le Cun, 2019).
4. Apprentissage-machine/ Apprentissage automatique	« Ensemble de méthodes pour entraîner un système, au lieu de le programmer explicitement. Dans l'apprentissage supervisé, le système est entraîné à accomplir une tâche à partir d'exemples d'entrée et de sortie correspondante. Dans l'apprentissage non supervisé ou autosupervisé, le système découvre les interdépendances entre les variables d'entrée sans être entraîné pour une tâche particulière. » (Le Cun, 2019).

5. Apprentissage profond « *L'apprentissage profond est une fonction d'intelligence artificielle (IA) qui imite le fonctionnement du cerveau humain dans le traitement des données et la création de modèles à utiliser dans la prise de décision.* » (Hargrave, 2021).

6. Back-up « Sauvegarde » en français.
« *La sauvegarde désigne le processus consistant à faire des copies de données ou de fichiers de données à utiliser en cas de perte ou de destruction des données ou des fichiers de données originaux.* » (Techopedia, 2016).

7. Big Data « *Le terme "big data" désigne les ensembles d'informations vastes et diversifiés qui se développent à un rythme toujours plus rapide. Il englobe le volume d'informations, la vélocité ou la vitesse à laquelle elles sont créées et collectées, ainsi que la variété ou la portée des points de données couverts.* » (Segal, 2021).

8. Business model « *Le terme business model fait référence au plan d'une entreprise pour réaliser des bénéfices. Il identifie les produits ou services que l'entreprise prévoit de vendre, le marché cible qu'elle a identifié et les dépenses prévues.* » (Kopp, 2020).

9. CLN Compréhension du langage naturel
« *La compréhension du langage naturel est une branche de l'intelligence artificielle qui utilise des logiciels informatiques pour comprendre des entrées sous forme de phrases en utilisant du texte ou de la parole.* » (TechTarget Contributor, 2021).

10. Cloud « *Cloud (ou Cloud computing) est un ensemble de technologies utilisée ensemble pour offrir un espace de stockage et de la puissance de calcul à différentes échelles, de*

la commodité à la haute performance, par l'intermédiaire d'un réseau public (Internet) ou dédié (ligne propriétaire). » (Dosquet et al., 2020, p.18).

11. CNIL
Commission Nationale de l'Informatique et des Libertés
« Dans l'univers numérique, la CNIL est le régulateur des données personnelles. C'est une autorité administrative indépendante qui accompagne les professionnels dans leur mise en conformité et aide les particuliers à maîtriser leurs données personnelles et exercer leurs droits. Elle analyse l'impact des innovations technologiques et des usages émergents sur la vie privée et les libertés. » (CNIL, s.d.).
12. CSA
Conseil Supérieur de l'Audiovisuel
« Le Conseil Supérieur de l'Audiovisuel est l'autorité administrative indépendante en charge de la régulation du secteur des médias audiovisuels en Fédération Wallonie-Bruxelles. » (CSA, s.d.).
13. Deep Learning
Voir apprentissage profond (n°5).
14. Digital nativity
« Nativité numérique » en français.
Il s'agit d'une personne qui est née à l'ère numérique.
15. Domotique
« La domotique regroupe toutes les techniques permettant de contrôler, de programmer et d'automatiser une habitation. Pour ce faire, elle use des domaines de l'électronique, de la télécommunication ainsi que des automatismes. » (Gallay, 2019).
16. Écosystème
« Un écosystème numérique est un groupe de ressources informatiques interconnectées qui peuvent fonctionner comme une unité. » (Brush, 2019).

17. Enceinte connectée « Une enceinte connectée ou haut-parleur intelligent est un type de haut-parleur sans fil à commande vocale, associé à un assistant personnel intelligent intégré, qui propose des actions interactives et des commandes à mains libres à l'aide d'un ou plusieurs mots-clés » (Wikipedia, 2021).
18. Endpointing « La détection du endpointing, ou point final de la parole, vise à déterminer les points de départ et d'arrivée du signal vocal. En général, la détection des points de fin de parole est adaptée au traitement frontal de la parole. » (Zhang et al., 2020).
19. GAFA Google, Amazon, Facebook et Apple
20. Génération en langage naturel En anglais “Natural Language Generation” (NLG)
 « Le NLG est le processus qui consiste à produire une réponse textuelle en langage humain à partir de certaines données d'entrée. Ce texte peut également être converti en format vocal grâce à des services de synthèse vocale. » (Kavlakoglu, 2020).
21. GOFAI Good Old-Fashioned Artificial Intelligence
 « Méthodes d'IA « classiques » basées sur la logique, les règles et les algorithmes de recherche, telles qu'elles se pratiquaient avant l'avènement de l'apprentissage-machine. » (Le Cun, 2019).
22. GPS Global Positioning System/ Système de positionnement mondial
 « Il s'agit d'un système mondial de navigation par satellite qui assure la synchronisation de la localisation, de la vitesse et du temps. » (Geotab, 2021).

23. IA Intelligence artificielle
« L'intelligence artificielle consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines d'imiter une forme d'intelligence réelle » (Futura Tech, s.d.)
24. IBM International Business Machines Corporation
« IBM est une entreprise leader dans le domaine des plateformes de cloud computing et des solutions cognitives. » (LinkedIn, s.d.).
25. IHM Interface Homme-Machine ou Interaction Homme-Machine
« Discipline s'occupant de la conception, de l'évaluation et de l'implémentation d'un système informatique interactif en vue de son usage par l'être humain et de l'étude des phénomènes majeurs qui l'accompagnent. L'objet principal d'étude de l'IHM est l'interface homme-machine, qui concerne les moyens par lesquels l'utilisateur final interagit avec un système, en particulier via des dispositifs d'interaction et du logiciel. » (Vanderdonckt, 2020)
26. Internet des Objets *« L'internet des objets (IdO) désigne un système d'objets interconnectés et reliés à internet, capables de collecter et de transférer des données sur un réseau sans fil sans intervention humaine. »* (Aeris, 2020).
27. iOS iPhone Operating System/ Système d'exploitation d'iPhone
« Système d'exploitation propre à la marque Apple et qui concerne donc toutes les gammes d'appareils de types iPhone, iPod, le iPad et les périphériques compatibles. » (Futura Tech, s.d.)

28. Linguistique computationnelle ou informatique *“La linguistique informatique se concentre sur l'application de méthodes quantitatives et statistiques pour comprendre comment les humains modélisent le langage, ainsi que sur des approches informatiques pour répondre aux questions linguistiques.”* (Kamath et al. 2019)
29. Machine Learning Voir apprentissage-machine (n°4).
30. MBUX Mercedes-Benz User Experience
« MBUX est une suite de technologies de pointe qui utilise l'intelligence artificielle pour améliorer chaque moment que vous passez sur la route dans votre nouvelle Mercedes-Benz. »
 (Mercedes-Benz, s.d.).
31. MIT Massachusetts Institute of Technology
« Le Massachusetts Institute of Technology (MIT) est une université privée de recherche située à Cambridge, dans le Massachusetts. Fondé en 1861, le MIT a depuis joué un rôle clé dans le développement de la technologie et de la science modernes, ce qui le classe parmi les institutions académiques les plus prestigieuses du monde. » (Wikipedia, 2021).
32. NLU Natural Language Understanding
 Voir CLN (n°9).
33. PDA Personal Digital Assistant/ Assistant numérique personnel
« L'assistant numérique personnel est un terme désignant un petit appareil mobile de poche qui offre des capacités de calcul, de stockage et de récupération d'informations pour un usage personnel ou professionnel, souvent pour garder à portée de main les horaires, les calendriers et les informations du carnet d'adresses. » (Steele, 2016).

34. Phonème *“Unité distinctive au sein des fréquences, exprimée en millième de seconde, permettant de distinguer les mots les uns des autres.” (Chapuzet, 2019)*
35. Phonographe *« Le phonographe est un appareil destiné d'abord à graver, puis à reproduire du son (paroles ou musiques) par un procédé purement mécanique. » (Wikipedia, 2021).*
36. Quality Data *« Quality data, ou qualité des données, est un terme utilisé par les professionnels de la gestion des données pour décrire une caractéristique des données qui peut être mesurée ou évaluée par rapport à des normes définies afin de déterminer la qualité des données. » (Janev et al., 2016).*
37. RGPD *Règlement général sur la protection des données*
« Le règlement général sur la protection des données (RGPD) est un cadre juridique qui fixe des lignes directrices pour la collecte et le traitement des informations personnelles des personnes qui vivent dans l'Union européenne (UE). » (Frankenfield, 2020).
38. STT *Speech-to-text*
« La reconnaissance vocale est la capacité d'une machine ou d'un programme à identifier les mots prononcés à haute voix et à les convertir en texte lisible. » (Kiwak, 2020).
39. Système embarqué *« Un système embarqué est un système électronique et informatique qui se trouve au sein d'un produit plus large. Son but premier consiste à traiter de l'information reçue depuis son environnement pour en faire profiter le produit hôte qui l'héberge. » (UKO, s.d.).*

40. TALN

Traitement Automatique du Langage Naturel

« Une discipline qui étudie la compréhension, la manipulation et la génération du langage naturel par les machines, que ce langage soit écrit ou parlé. Il allie les connaissances issues de la linguistique, de l'informatique et de l'intelligence artificielle. » (Samain, 2018).

41. TTS

Text-to-Speech

« La conversion du texte en parole est un processus de modélisation du langage naturel qui nécessite de transformer des unités de texte en unités de parole pour une présentation audio. C'est l'inverse de la conversion de la parole en texte (speech-to-text). » (Techopedia, 2012).

ANNEXE 2 : Hype cycle de Gartner incluant les assistants virtuels

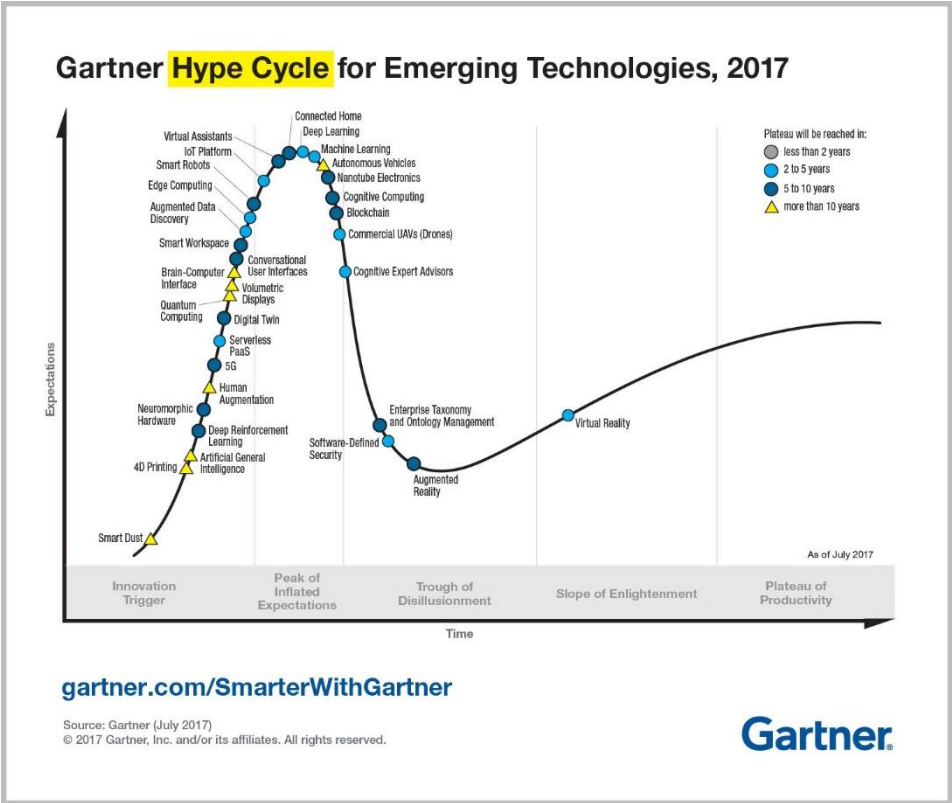


Figure 10: Courbe de Gartner pour les technologies émergentes en 2017
(Source : Gartner, 2017)

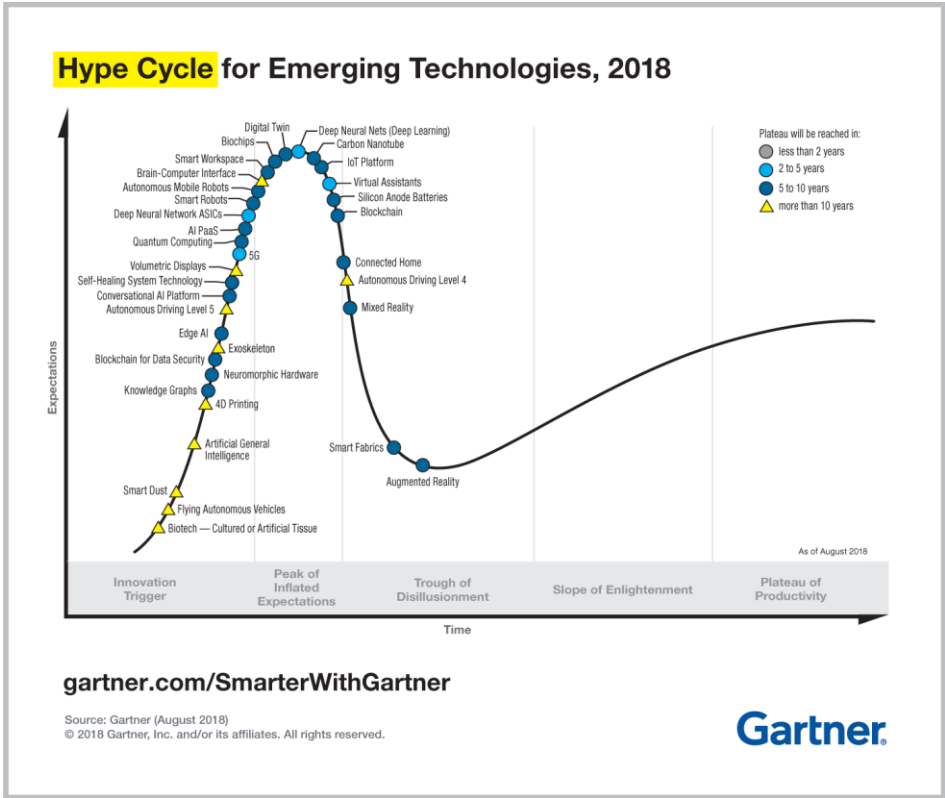


Figure 11: Courbe de Gartner pour les technologies émergentes en 2018
(Source : Gartner, 2018)

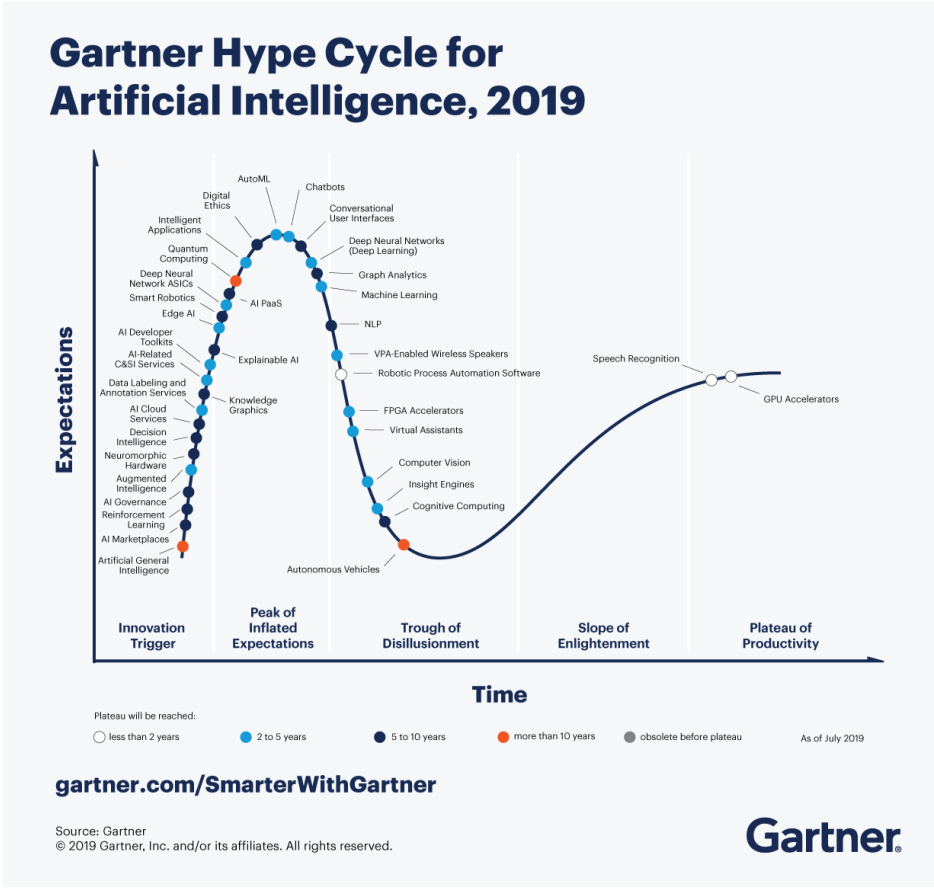


Figure 12: Courbe de Gartner pour l'intelligence artificielle en 2019
(Source : Gartner, 2019)

ANNEXE 3 : Experts contactés

	Personnes contactées	Profils	Réponse
Cas d'usage n°1 : Aide à la conduite	1. Amaury Delloye	Membre du conseil d'administration à adn.ai, plateforme de publicité conversationnelle	Pas de réponse
	2. Jean-François Bonastre	Professeur au Laboratoire d'Informatique d'Avignon et spécialiste du traitement de la parole et de l'authentification vocale	Positive
	3. Nicolas Obin	- Professeur associé à l'Université de la Sorbonne, en Technologies de la parole, de l'audio et de la musique - Chercheur dans l'équipe d'analyse et de synthèse sonore de l'Ircam - Centre Pompidou	Pas de réponse
	4. Thomas Drugman	- Scientifique ayant une expérience en apprentissage automatique et en traitement du signal - Responsable scientifique de l'apprentissage automatique à Amazon	Pas de réponse
	5. Susana Sánchez Restrepo	Doctorante en robotique et automatique avec une formation en mécatronique	Pas de réponse
	6. Michael Hansen	Informaticien au laboratoire de recherche de l'armée de l'air (Air Force Research Laboratory, Ohio)	Pas de réponse
	7. Koen Vervloesem	Journaliste informatique et scientifique	Pas de réponse
	8. Preston So	Auteur de « Voice Content and Usability »	Pas de réponse
	9. Jon Stine	Directeur exécutif à « The Open Voice Network » (Portland)	Pas de réponse
	10. Eric Dosquet	Auteur de plusieurs livres sur l'innovation dont « <i>La Révolution des Assistants Vocaux - Comprendre les enjeux et réussir ses stratégies marketing</i> »	Suggère de contacter Oxana Gouliaéva

11.	Oxana Gouliaéva	Auteur du livre “ <i>La Révolution des Assistants Vocaux - Comprendre les enjeux et réussir ses stratégies marketing</i> ”	Positive
12.	Yvon Moysan	Directeur Académique Master Digital Marketing et Innovation IESEG School of Management	Pas de réponse
13.	Ariane Nabeth-Halber	- Directrice “Speech Solutions” chez BertinIT - Chercheuse en reconnaissance de parole dans les laboratoires Ater (Japon) ou Thales (France)	Positive
14.	Olivia Longo	Intelligent Assistant Content Manager à Inago, une plateforme d'assistant IA conversationnel pour les produits	Pas de réponse
15.	Ron DiCarlantonio	Fondateur & CEO de Inago	Pas de réponse
16.	Nicolas Perony	Co-fondateur de OTO.ai (New-York), entreprise de logiciels informatiques	Pas de réponse
17.	Teo Borschberg	Co-fondateur & CEO de OTO.ai	Pas de réponse
18.	Richard Lepine	Professionel en Marketing Digital Automobile	Pas de réponse
19.	Soledad Bedu	Ingénieur linguiste & Voice user interface designer chez Faurecia Automotive, spécialisée dans l’UX conversationnel	Positive
20.	Robin Guérit	Co-fondateur de Flowchase	Positive
21.	Zoé Broisson	Co-fondatrice de Flowchase	Pas de réponse
22.	Ludovic Chevalier	Co-fondateur de Speaky, un réseau social pour apprendre des langues en pratiquant avec des personnes du monde entier	Pas de réponse
23.	François-Xavier Fievez	- Expert spécialisé dans le domaine de la didactique des langues étrangères et détaché pédagogique et professeur d’anglais à ELV (Ecole des Langues Vivantes), UNamur	Positive
24.	Justine Esteve	Spécialiste de la formation en langues	Pas de réponse
25.	Maaike Groenewege	Linguiste conversationnelle	Pas de réponse
26.	Antoine Besnehard	Responsable de l'innovation pédagogique at Fondation I-SITE Université Lille Nord-Europe (ULNE)	Suggère de contacter Luigi Lancieri
27.	Aline Chauvier	Traductrice	Pas le profil approprié

Cas d'usage n°2 : Apprentissage d'une langue étrangère	28.	Philippe Deltenre	Consultant spécialisé dans l'innovation d'IA	Pas le profil approprié
	29.	Nicolas Roland	Learning Experience Designer	Pas de réponse
	30.	David Moussebois	Senior Manager Learning Development	Pas de réponse
	31.	Luigi Lancieri	Professeur de sciences informatiques à l'Université de Lille et en recherche au Laboratoire D'informatique fondamentale de Lille (LIFL)	Pas de réponse
	32.	Mónica Pamela Acosta Rodas	- Professeur et chercheuse à l'Universidad Catolica del Ecuador - Auteur de plusieurs articles scientifiques sur l'apprentissage virtuel	Pas de réponse
	33.	Omar Cándor-Herrera	- Auteur de plusieurs articles scientifiques sur l'apprentissage virtuel et l'intelligence artificielle	Pas de réponse
	34.	Bertrand Jouvenot	Président & fondateur des « brigades du marketing »	Suggère de contacter Maria Perillo et Brian Subirana
	35.	Maria Perillo	- Chercheuse au MIT et a développé le premier test du niveau d'anglais avec Alexa - Learning Experience Design (LXD) à IESE Business School (Barcelone)	Positive
	36.	Brian Subirana	- Chercheur au MIT et a développé le premier test du niveau d'anglais avec Alexa	Pas de réponse
	37.	Vu van	Co-fondatrice & CEO de ELSA (English Language Speech Assistant), l'assistant I.A. le plus intelligent au monde pour l'apprentissage de la prononciation et la réduction des accents (concurrent de Flowchase)	Pas de réponse
	38.	Marielle Henriët	Maître de langues à l'UCLouvain	Suggère de contacter Zoé Broisson

Cas d'usage n°3 : Services de secours à distance	39.	Manuel Kolpe	Professeur en Analyse et conception de systèmes d'information à l'UCLouvain	Pas de réponse
	40.	Ahmed Bouzid	- Fondateur et CEO de Witlingo (Washington) - Auteur collaborateur de Voicebot.ai - Responsable produit chez Amazon	Positive
	41.	Etienne Riviere	- Professeur en ingénierie des systèmes et cyber sécurité à l'UCLouvain	Suggère de contacter Stanley Goffinet
	42.	Axel Legay	- Professeur en ingénierie des systèmes et cyber sécurité à l'UCLouvain - Auparavant professeur à l'université de Carnegie Mellon (Pittsburgh) et professeur à l'université de Californie (Santa Cruz)	Positive
	43.	Christophe Tricot	Co-fondateur & CEO de la Forge, un Accélérateur dédié à l'Intelligence Artificielle	Pas de réponse
	44.	Stanley Goffinet	- Étudiant en master 2 en sciences informatiques à l'école polytechnique de Louvain - Mémoire publié: " <i>Controlling security rules using natural dialogue: An application to smart home care</i> "	Positive
	45.	Roger Moore	Chair of Spoken Language Processing à l'Université de Sheffield	Pas de réponse

ANNEXE 4 : Guides d'entretien

Lettre d'introduction

Madame, Monsieur,

Dans le cadre de nos études universitaires en sciences de gestion, nous sommes actuellement en train de rédiger notre mémoire au sujet des assistants vocaux. L'objectif de celui-ci est d'analyser 3 cas pratiques dans lesquels l'assistant vocal apporterait une plus-value. Ces cas dont il est question sont l'aide à la conduite, l'apprentissage d'une langue étrangère ainsi que les services de secours à distance.

Pour ce faire, nous sommes à la recherche d'experts à interviewer, qui pourraient répondre à quelques questions lors d'un entretien en visioconférence (entre 30 et 60 minutes). C'est pourquoi, nous nous permettons de vous contacter car nous souhaiterions aborder cette thématique avec vous. Nous serions ravies de vous adresser ces questions et cela nous serait d'une grande aide.

Seriez-vous disposé à participer à cette interview ?

D'avance merci,

Bien à vous,

Alice & Pauline

Cas d'usage n°1 : Aide à la conduite

Introduction

Bonjour, merci d'avoir accepté de participer à cet interview. Pour rappel, l'objectif de notre mémoire est d'analyser 3 cas d'usage dans lesquels l'assistant vocal apporterait une plus-value. Le cas qui sera étudié au cours de l'entretien n'est autre que l'aide à la conduite. La durée de cette interview n'excédera pas 1 heure. Les questions seront adressées par ordre de priorité afin de respecter le timing.

Notre mémoire vise à comprendre l'influence de **l'utilisation de l'assistant vocal dans les voitures**. Étant donné que la conduite réclame une attention particulière, il est préférable de ne pas quitter la route des yeux pour effectuer différentes tâches dans sa voiture. Mais qu'en est-il de la réelle plus-value apportée par un assistant vocal dans l'aide à la conduite ?

Au travers de cet entretien, nous souhaiterions mieux appréhender comment l'assistance vocale et la façon dont elle peut être utilisée dans les voitures.

Présentation de la personne interviewée

1. Pourriez-vous vous présenter en quelques mots ? Quel est votre parcours ?
2. Quel est votre métier/ rôle dans le domaine de l'intelligence artificielle ?
3. Quelle est votre expertise en matière d'assistants vocaux

« Warm-up » assistants vocaux

4. Selon vous, quelles sont les avancées actuelles des voitures dotées d'une assistance vocale ?
5. Selon vous, l'assistance vocale apporte-t-elle une plus-value aux voitures qui en sont équipées ?

Focus

6. Pensez-vous que l'assistant vocal au sein des voitures peut apporter une aide ou une sécurité supplémentaire à la conduite ?

7. Considérez-vous les assistants vocaux dont les voitures sont dotées comme un gadget ou plutôt comme une innovation indispensable ?
8. Les nouvelles technologies qui arrivent sur le marché réclament toujours un temps d'adaptation. Où se situe-t-on par rapport en termes d'adoption d'assistant vocal dans les voitures ?
9. Certains automobilistes sont toujours réfractaires à l'idée d'utiliser la commande vocale ? Quelles seraient selon vous les freins les plus importants parmi la nécessité d'une diction précise, la duplication des commandes, ... ?
10. Pensez-vous que les constructeurs automobiles ont intérêt à intégrer leur propre assistant vocal dans leur habitacle ou est-ce que les utilisateurs préfèrent utiliser l'assistant vocal de leur smartphone ? Pensez-vous que l'automobiliste préfère interagir de façon spécifique avec sa voiture ou est-ce que les fonctions que lui offre l'assistant vocal lui suffisent ?
11. La plus grosse crainte encourue à l'égard de l'utilisation d'un assistant vocal concerne la gestion de leurs données personnelles, qu'en pensez-vous ?
12. Quelles sont les demandes qui restent jusqu'à lors trop complexes à traiter pour un assistant vocal ?
13. À l'avenir, pensez-vous que cela sera la norme que les constructeurs équipent tous les véhicules d'assistants vocaux ?
14. Si cela venait à se produire, serait-il plus opportun pour les constructeurs d'intégrer des assistants vocaux de Google, Amazon ou Apple par exemple ou bien de se doter de leur propre technologie ? Dans le premier cas, cela contribuerait-il davantage au monopole exercé par les GAFA ?

Discussions sur les prévisions futures

15. Pensez-vous que lorsque l'assistant vocal sera capable de reconnaître les émotions dans la voix, cela sera davantage utile dans l'assistance à la conduite ?
16. Selon vous, quel est le futur des assistants vocaux au sein des voitures ?

Approfondissement & Conclusion

Résumé de l'avis, clarifications si besoin et remerciements

Cas d'usage n°2 : Apprentissage d'une langue étrangère

Introduction

Bonjour, merci d'avoir accepté de participer à cet interview. Pour rappel, l'objectif de notre mémoire est d'analyser 3 cas d'usage dans lesquels l'assistance vocale apporterait une plus-value. Le cas qui sera étudié au cours de l'entretien n'est autre que l'apprentissage d'une langue. La durée de cette interview n'excédera pas 1 heure. Les questions seront adressées par ordre de priorité afin de respecter le timing.

Notre mémoire vise à comprendre l'influence de l'utilisation de **l'assistance vocale dans l'apprentissage d'une langue étrangère**. Étant donné que l'apprentissage d'une langue réclame une précision et une assiduité particulière, la question de comment améliorer leur efficacité et leur faciliter la tâche se pose. Alors qu'en est-il de la réelle plus-value apportée par un assistant vocal en tant qu'outil pédagogique ?

Au travers de cet entretien, nous souhaiterions mieux appréhender comment l'assistance vocale peut apporter une plus-value dans l'apprentissage d'une langue étrangère.

Présentation de la personne interviewée

1. Pourriez-vous vous présenter en quelques mots ? Quel est votre parcours ?
2. Quel est votre métier/ rôle dans le domaine de l'intelligence artificielle ?
3. Quelle est votre expertise en matière d'assistants vocaux ?

« Warm-up » assistants vocaux

4. Connaissez-vous les avancées actuelles mises en place en termes d'assistants vocaux qui peuvent assister dans l'apprentissage d'une langue ?
5. Dans le cadre de l'apprentissage d'une langue, pensez-vous que l'assistant vocal peut apporter une aide significative ? Est-ce que ça rendrait l'apprentissage d'une langue plus qualitatif et/ou plus rapide ?

Focus

6. Dans quelle compétence est-ce que l'usage de l'assistant vocal sera utile : orale/ écrite/ audition/ lecture ? que pour la compétence prononciation ?
7. Selon vous, où se positionnent les enseignants et les élèves en termes d'adoption d'assistance vocale ?
8. Est-ce que l'assistant vocal pourrait également être utile dans le cadre d'un apprentissage autodidacte ?
9. L'apprentissage d'une langue à l'oral se fera-t-il à l'aide d'assistants vocaux, via des enceintes connectées développées, par exemple, par Google ou Amazon ou en faisant appel à des applications conçues spécifiquement à cet effet ? Quelle serait la meilleure option ?
10. Du point de vue des élèves, quels peuvent être les freins à l'idée d'utiliser un assistant dans l'apprentissage d'une langue ? Qu'en est-il du point de vue des enseignants ?
11. Selon vous, est-ce que l'utilisation de l'assistance vocale dans ce cas d'usage pourrait constituer un quelconque danger ?
12. Pensez-vous que l'assistant vocal sera plus utile pour assister les étudiants que les professeurs de langue ?
13. Pensez-vous que l'assistant vocal jouera le rôle de complémentarité à l'enseignement ?

Discussions sur les prévisions futures

14. Pensez-vous que cette forme d'intelligence pourrait remplacer l'enseignement ?
15. Comment voyez-vous le futur de l'enseignement des langues ? Pensez-vous que les assistants vocaux seront davantage utilisés par les enseignants et aideront les élèves ?

Approfondissement & Conclusion

Résumé de l'avis, clarifications si besoin et remerciements

Cas d'usage n°3 : Services de secours à distance

Introduction

Bonjour, merci d'avoir accepté de participer à cet interview. Pour rappel, l'objectif de notre mémoire est d'analyser 3 cas d'usage dans lesquels l'assistance vocale apporterait une plus-value. Le cas qui sera étudié au cours de l'entretien n'est autre que les services de secours à distance. La durée de cette interview n'excédera pas 1 heure. Les questions seront adressées par ordre de priorité afin de respecter le timing.

Notre mémoire vise à comprendre l'influence de **l'utilisation de l'assistance vocale dans les situations d'urgence à domicile**, en particulier pour les seniors ou les personnes vulnérables, qui ne font pas preuve d'autonomie. Étant donné que la demande de secours en cas d'urgence réclame une certaine rapidité et des informations précises, il serait opportun d'avoir un moyen efficace à disposition afin de demander de l'aide instantanée. Mais qu'en est-il de la réelle plus-value apportée par un assistant vocal dans l'aide aux personnes, à distance ?

Au travers de cet entretien, nous souhaiterions mieux appréhender comment l'assistance vocale utilisée au travers d'enceintes connectées permettrait de renforcer la sécurité et de rendre les services de secours plus efficaces.

Présentation de la personne interviewée

1. Pourriez-vous vous présenter en quelques mots ? Quel est votre parcours ?
2. Quel est votre métier/ rôle dans le domaine de l'intelligence artificielle ?
3. Quelle est votre expertise en matière d'assistants vocaux ?

« Warm-up » assistants vocaux

1. Selon vous, quelles sont les avancées actuelles dans les assistants vocaux dédiés aux secours à distance ?
2. Selon vous, l'assistance vocale apporte-t-elle une plus-value pour les services de secours à distance ?

Focus

3. Pensez-vous que les personnes âgées/ vulnérables pourraient demander de l'aide et contacter les secours plus rapidement en s'adressant à un assistant vocal ? La voix n'est pas forcément le meilleur moyen ?
4. Selon vous, est-ce que la présence d'un assistant vocal permettra aux personnes de se sentir plus en sécurité ? De la même manière, est-ce que cela pourra rassurer de savoir ses proches en possession d'un assistant vocal ?
5. Est-ce que la présence d'un assistant vocal pourrait augmenter l'autonomie des seniors ou des personnes en situation de handicap ?
6. Selon vous, quels sont les risques d'implémenter un assistant vocal dans une situation d'urgence ?
7. Selon-vous, où se situe-t-on en termes d'adoption des assistants vocaux dans les situations d'urgence ? Pensez-vous que les personnes en détresse sont prêtes à faire confiance à un assistant vocal plutôt qu'à une personne ?
8. Pensez-vous que l'adoption d'une enceinte connectée est suffisamment implémentée afin de l'utiliser en cas d'urgence ?
9. Est-ce que la barrière technologique constitue un frein à l'utilisation d'un assistant vocal pour les personnes âgées/ personnes vulnérables ?
10. L'aide à distance à l'oral se fera-t-elle à l'aide d'assistants vocaux, via des enceintes connectées développées, par exemple, par Google ou Amazon ou en faisant appel à des applications conçues spécifiquement à cet effet ? Quelle serait la meilleure option ?
11. Quelles sont les demandes encore trop complexes à traiter pour un assistant vocal dans des situations d'urgence à distance ?

Discussions sur les prévisions futures

12. Pensez-vous que lorsque l'assistant vocal sera capable de reconnaître les émotions dans la voix, cela sera une plus-value dans l'assistance aux services d'aide à distance ?
13. Que pensez-vous être le futur des services de secours à distance aidés par les assistants vocaux ?

Approfondissement & Conclusion

Résumé de l'avis, clarifications si besoin et remerciements

ANNEXE 5 : Matrice de synthétisation des interviews

Cas d'usage n°1 : Aide à la conduite

	Expert 1 : Oxana Gouliaéva	Expert 2 : Ariane Nabeth-Halber	Expert 3 : Jean-François Bonastre	Expert 4 : Soledad Bedu
Plus-value	Répond à certaines contraintes : le gain de temps et les mains occupées.	Ça peut apporter une sécurité.	Représente une aide pour des petits usages, mais pour des vrais usages il y a de gros problèmes d'ergonomie et de performance.	Si l'assistant vocal est bien conçu et intuitif pour l'utilisateur, il représentera une plus-value. L'assistant assure une certaine rapidité et facilité. L'assistant aide le conducteur à avoir une conduite plus sécuritaire, plus confortable et plus personnalisée.
Apporte une sécurité ?	Pas forcément une sécurité en plus. Cela dépend des usages ainsi que de la performance de l'assistant vocal.	Tout ce qui peut permettre de faire des requêtes à la voix sans enlever ses yeux de la route, c'est intéressant.	Oui pour les actions comme téléphoner, taper un message à l'écran, ... Mais pas pour l'aide à la conduite en elle-même. Il peut y avoir un avantage sécuritaire si l'assistant reconnaît l'état émotif du conducteur.	Là où c'est sécuritaire, c'est que l'utilisateur peut faire des choses sans avoir à enlever les mains du volant ou quitter la route des yeux. Les assistants vocaux proactifs peuvent également apporter une sécurité.

				La détection de la fatigue → l'assistant vocal peut agir.
Risques	Le fait de discuter dans la voiture nous distrait → Il y a quand même une charge cognitive importante.	Il y a également le risque que ça peut détourner l'attention si la charge cognitive est trop importante. Système de <i>endpointing</i> → tâche qui ne demande aucune charge cognitive, mais c'est devenu un effort, qui demande une grande concentration.	Risque que la communication occupe une partie de la capacité cognitive. Risque que l'assistant vocal trace en permanence l'utilisateur en termes GPS.	Risque d'avoir une charge cognitive trop importante qui ne sera pas bonne. Risque que l'assistant soit envahissant en allant au-devant de tous les besoins
Facteurs favorisant l'adoption	Performance : compréhension de la requête et pertinence de réponse ou de l'action. L'expérience offerte dans la voiture plutôt que la qualité de la conduite.	Si le business model est bien fait et compatible avec les attentes de l'utilisateur.	/	Il est nécessaire que la voix soit plus simple que le toucher. Avoir un dialogue sans avoir à regarder l'écran. Avoir un assistant vocal facile et pratique (comme Android Auto et CarPlay).
Freins à l'adoption	Critères de choix et de préférences de l'utilisateur (design, expérience). Cela dépend aussi du <i>digital nativity</i> et également de la région du monde dans laquelle l'utilisateur se trouve.	Barrières à l'entrée : règles culturelles, dialogiques, métiers dans toutes les langues, et non la technologie de reconnaissance vocale. Si le business model n'est pas compatible avec la demande de	Limitations techniques et manque d'intérêt, par exemple de savoir si l'utilisation de la voix est réellement plus pratique. Il y a également un manque d'applications ciblées à l'heure actuelle.	Ne pas savoir ce comment les données personnelles sont utilisées (notamment avec Google et Amazon). Le fait que l'assistant écoute en permanence, il y a une méfiance quant aux conversations privées.

		l'utilisateur, l'assistant vocal ne sera pas adopté.	Il y a aussi des réticences sur la préservation des informations personnelles.	Il y a la crainte de se tromper et d'échouer. Le fait de ne pas être compris car on n'a pas le bon accent.
Technologie embarquée ou externe ?	Les constructeurs automobiles cherchent à intégrer l'assistance vocale en natif.	C'est sujet à débat mais exemple avec Faurecia : Complémentarité entre technologie embarquée et technologie externe.	Technologie externe. Il y a un gros conflit entre le système d'information des voitures et l'assistant vocal sur la plupart des voitures, l'assistant vocal embarqué n'évolue pas. Technologie embarquée pour la partie acoustique (neutralise le bruit) mais technologie externe pour la partie application.	Complémentarité entre le Cloud et la technologie embarquée. Si perte de 4G, la technologie embarquée ne répond plus → système de backup avec le cloud → mieux d'avoir les deux. Pense que le cloud permet une meilleure reconnaissance.
Technologie développée par les constructeurs automobiles ou autres ?	Les constructeurs de voitures ont tout intérêt à développer leurs assistants afin de conserver une relation directe avec le conducteur.	Exemple avec Faurecia : Complémentarité entre Alexa et l'assistant vocal interne à la voiture.	Les assistants vocaux peuvent être définis de manière transversale aux marques. Des acteurs spécialisés de la voiture vont se développer.	Cela dépend des moyens financiers des constructeurs → s'ils savent investir ils peuvent créer un assistant vocal embarqué. Cela dépend aussi de la clientèle → clientèle âgée, familles. Cela dépend également de l'image de marque.
Fonctionnalités spécifiques à la voiture ou	C'est intéressant de pouvoir accéder à plus de services	Complémentarité entre les deux. Technologie embarquée pour les fonctionnalités propres à la	/	Fonctionnalités propres à la voiture tant qu'on n'est pas en conduite autonome.

fonctionnalités générales ?	différents qui ne concernent pas que la voiture.	conduite et technologie externe pour les fonctions génériques.		
Prédictions futures	Interaction avec plusieurs assistants.	/	Peut-être régulateur de vitesse mais il n'est pas convaincu de cette fonctionnalité. Vers une gamme de services plus restreints et renforcés, qui seront très liés à la conduite. L'authentification par la voix : voix en tant que mot de passe.	Si voiture autonome → avoir un assistant vocal pas seulement pour la conduite et pas seulement pour le conducteur. Avoir des conversations plus humaines mais comprendre l'ironie, par exemple, va être très compliqué. Détection des émotions par la voix et sur le visage. Avoir un assistant vocal dans la voiture va devenir un standard. Vers des assistants comme MBUX.

Cas d'usage n°2 : Apprentissage d'une langue étrangère

	Expert 1 : François-Xavier Fievez	Expert 2 : Robin Guérit	Expert 3 : Maria Perillo
Plus-value	Aide à l'apprentissage et plus qualitatif → possibilité de s'entraîner seul et de pratiquer autant de fois qu'on veut. C'est également plus rapide et permet une certaine autonomie. Pédagogie à distance → avoir une présence à distance qui encadre l'étudiant.	Donne des feedbacks personnalisés avec des objectifs à atteindre. Aide aux professeurs.	Plus d'exposition → possibilité d'utiliser Alexa en faisant autre chose et être exposé à la compréhension orale. Apporte une aide à l'apprentissage → possibilité d'écouter, d'imiter l'intonation et la prononciation. Améliore la prononciation, surtout en anglais. Timidité → les personnes osent plus parler à un assistant vocal qu'à un humain car pas de jugement.
Facteurs favorisant l'adoption	Si l'utilisation est instinctive. Si ça permet de faire quelque chose qu'on ne saurait pas faire autrement.	Sera adopté si c'est en complément à une formation. Il faut qu'y ait un feedback personnalisé et pas juste de la comparaison. Il faut que l'assistant vocal agisse en tant que cadre pédagogique et pas seulement en tant que divertissement → voir qu'on peut vraiment apprendre grâce à l'assistant.	Le design de la plateforme. Exemple avec Alexa et possibilité de créer des skills. Si Alexa déjà présent dans les domiciles → favorise l'adoption car plus d'exposition.

Freins à l'adoption	Résistance de la part des professeurs de changer de méthode de travail et crainte que ça puisse les remplacer. Si on n'en voit pas la réelle utilisation et que ça apporte une charge supplémentaire. Peur de ne pas savoir maîtriser l'assistant vocal. Le coût de l'assistant vocal/ application. L'utilisation des données personnelles.	Pas encore assez développé pour que ça fonctionne à 100% pour tout. Les différents accents → pas encore assez de données pour accepter tous les accents des non-natifs.	Si l'élève est débutant, Alexa ne le comprendra pas → parler à un professeur en premier. <i>Endpointing</i> et sémantique (5% non atteint correspondant à l'ambiguïté). Situation Covid → beaucoup adoptent une approche hybride avec l'e-learning. Pas encore bien adopté en Europe. Barrière de ne pas savoir utiliser l'assistant. Pas possible de modifier la langue de l'assistant vocal.
Assistant vocal comme outil pédagogique ?	C'est plus utile dans le cadre d'apprentissage autodidacte, ou en plus des cours pour pratiquer seul.	Oui, il devrait être validé pédagogiquement par des institutions qui sont renommées, comme les universités.	/
Complémentaire au professeur ou application indépendante ?	Il faut une complémentarité de la tâche du professeur, une pratique autonome avec l'assistant vocal.	Complémentaire au rôle du professeur avec des objectifs précis.	Complémentaire : il faut avoir une combinaison de plusieurs propositions → combinaison optimale = auto-apprentissage et contact humain. Exemple avec expérience au MIT avec Alexa.
Technologie développée par qui ?	Startups où il y a une association d'un spécialiste de la technologie et d'un spécialiste de la langue (comme	Par des startups. Les géants du web ne savent pas donner de feedbacks. Ils ne	Alexa s'améliore. Il faut essayer de concevoir un système éducatif en tenant compte des limites d'Alexa.

	Flowchase) → Avantage de pouvoir contacter les dirigeants et d'avoir une connaissance de l'utilisation des données personnelles, ce qui ne serait pas le cas avec des gros acteurs.	créent pas spécifiquement un outil pédagogique.	
Prédictions futures	Ça ne remplacera jamais l'humain. Ça ne prendra pas plus de place au sens péjoratif du terme.	La correction, évaluer le niveau, trouver des erreurs grammaticales, de sens, d'expressions, pourront se faire par la voix. Personnalisation : s'adapter à chaque secteur et pouvoir détecter les erreurs de chaque étudiant. Exemple Clubhouse : app où on verra un prof d'anglais qui est en fait un robot vocal.	Amélioration de sémantique afin de comprendre l'ambiguïté. Ne remplacera jamais le professeur. Mais les enseignants qui utilisent Alexa peuvent potentiellement remplacer ceux qui ne l'utilisent pas. Apprentissage renforcé car il y a plus d'exposition. Amélioration du problème de configuration de la langue de l'appareil.

Cas d'usage n°3 : Services de secours à distance

	Expert 1 : Axel Legay	Expert 2 : Ahmed Bouzid	Expert 3 : Stanley Goffinet
Plus-value	Augmente l'autonomie des personnes et leur sert de compagnie. Transmission automatique des données médicales au médecin/ aides-soignantes. Peut effectuer des tâches à leur place, peut leur faciliter la vie et leur diminuer des efforts.	Apporte une compagnie (se sentent moins seules, moins isolées). Permet d'être plus autonomes. Représente une facilité pour les proches qui ne doivent plus forcément être présents en permanence.	Représente une vraie aide et simple à utiliser. Le fait que ce soit personnalisable pour chaque utilisateur représente une plus-value. Augmente l'autonomie → rend le domicile indépendant et apte pour le vieillissement. Peut contrer la solitude tout en restant un complément à une personne physique.
Augmente la sécurité ?	Oui, ça va surtout faciliter le dialogue. Ça va renforcer la sécurité si l'assistant vocal est complémentaire à l'être humain. Ça peut également être rassurant pour les proches. Ça rend les choses plus confortables (identifie chute de gravité, la non prise de pilule, etc).	Oui, c'est plus sûr que s'il n'y a pas d'assistant vocal → apporte quelque chose de supplémentaire.	Oui ça augmente la sécurité et c'est également rassurant pour les proches.
Risques	Risque de mauvais fonctionnement. Risque de <i>hackage</i>.	Importantes conséquences de l'échec si le design n'est pas bien fait.	Risque que l'assistant ne comprenne pas.

	Risque d'investir dans l'assistant et qu'il ne soit pas adopté.	Risque que l'action ne se réalise pas → si réaction trop lente, si le Wifi ne fonctionne plus ou s'il n'appelle pas les secours quand il y a la nécessité. Risque de ne pas savoir qui écouter si plusieurs personnes dans la pièce. Risque d'ambiguïté → comment savoir s'il s'agit vraiment d'une urgence ? Risque du faux positif → ça appelle les secours mais rien de grave ne s'est passé.	Risque qu'il n'y ait plus de connexion internet.
Facteurs favorisant l'adoption	Effort pédagogique. Personnalisable (personnes de contact et façon d'interagir). Transparence sur la façon dont les données privées sont gérées.	Nécessaire que l'application soit bien faite car il s'agit d'une situation délicate. Avoir une connaissance de la façon dont ça affecte les services de secours, les utilisateurs et les proches. Savoir comment l'utiliser.	Il faut que ça soit personnalisable → gestionnaire pour paramétrer en fonction de la personne. L'assistant vocal ne doit pas remplacer les anciens systèmes → complémentaire Éduquer les personnes âgées pour l'utilisation OU rendre le système tellement simple que c'est inné.
Freins à l'adoption	Intrusion dans la vie personnelle. Barrière technologique → peur de ne pas savoir s'en servir. Barrière pédagogique → manque d'explication.	Barrière technologique. <i>Endpointing</i> : pour début et fin → Début : pense avoir entendu le <i>trigger word</i> alors que non. Fin : pas possible de faire des pauses sans que l'assistant pense que la requête est finie.	Crainte par rapport à la technologie → peur de ne pas savoir l'utiliser. Complicé et coûteux d'installation lorsque la personne n'a pas de wifi ni téléphone.

		L'assistant vocal ne permet pas de parler plus d'un certain temps.	
Technologie développée par qui ?	Petites startups belges : gens qui vont sur place, avec des <i>clouds</i> souverains, sont transparents → plus cher mais plus sûr pour les données personnelles. Les grands acteurs enregistrent tout pour améliorer la reconnaissance vocale.	Par des entreprises spécifiques → ils iront loin dans la spécialisation du cas d'usage (bcp de recherche, comprennent le problème et sont engagés sur le long terme) → meilleure expérience. Exemple Amazon : ne cherche pas à se spécialiser mais à toucher le plus de monde possible (<i>business model</i>).	Cela dépend de la demande des utilisateurs → si aucune crainte par rapport aux données personnelles, n'importe quel système peut être utilisé. Sinon entreprise locale.
Prédictions futures	Être proactif et donner des renseignements aux services de secours. Assistants vocaux reliés à des puces dans votre corps et qui peuvent mesurer un certain nombre de choses. Ne remplacera pas certains métiers mais va plutôt en créer dans ce cas d'usage-ci.	Les assistants vocaux seront améliorés sur les points faibles en priorité (<i>endpointing</i>, reconnaissance vocale). Ils seront de plus en plus proactifs.	Les personnes seront de plus en plus enclines à adopter les assistants vocaux et installer le nécessaire dans le domicile. L'assistant vocal pourra traiter de plus en plus de cas de figure. Pourra capter le sentiment/ émotion et le refléter. Assistant vocal sur n'importe quel support (enceinte, téléphone, oreillette, frigo). Système domotique et pas seulement enceinte connectée par exemple.

ANNEXE 6: Transcription des interviews d'experts

Cas d'usage n°1 : Assistant vocal dans les voitures			
Profil de la personne interviewée		Date de l'interview	Page
1.	<u>Oxana Gouliaéva</u> <i>Fonction : Auteur du livre "La Révolution des Assistants Vocaux - Comprendre les enjeux et réussir ses stratégies marketing"</i>	22 juin 2021	148
2.	<u>Ariane Nabeth-Halber</u> <i>Fonction :</i> - Directrice "Speech Solutions" chez BertinIT - Chercheuse en reconnaissance de parole dans les laboratoires Ater (Japon) ou Thales (France)	23 juin 2021	158
3.	<u>Jean-François Bonastre</u> <i>Fonction : Professeur au Laboratoire d'Informatique d'Avignon et spécialiste du traitement de la parole et de l'authentification vocale</i>	28 juin 2021	174
4.	<u>Soledad Bedu</u> <i>Fonction : Ingénieur linguiste & Voice user interface designer chez Faurecia Automotive, spécialisée dans l'UX conversationnel</i>	12 juillet 2021	185
Cas d'usage n°2: Apprentissage d'une langue étrangère			
5.	<u>François-Xavier Fievez</u> <i>Fonction :</i> - Expert spécialisé dans le domaine de la didactique des langues étrangères - Détaché pédagogique et professeur d'anglais à ELV, UNamur	7 juillet 2021	202
6.	<u>Robin Guérit</u> <i>Fonction : Co-fondateur de Flowchase</i>	9 juillet 2021	214

7.	<u>Maria Perillo</u> <u>Fonction :</u> - Chercheuse au MIT et a développé le premier test du niveau d'anglais avec Alexa - Learning Experience Design (LXD) à IESE Business School (Barcelone)	13 juillet	228
Cas d'usage n°3: Services de secours à distance			
8.	<u>Axel Legay</u> <u>Fonction :</u> - Professeur en ingénierie des systèmes et cyber sécurité à l'UCLouvain - Auparavant professeur à l'université de Carnegie Mellon (Pittsburgh) et professeur à l'université de Californie (Santa Cruz)	5 juillet 2021	238
9.	<u>Ahmed Bouzid</u> <u>Fonction :</u> - Fondateur et CEO de Witlingo (Washington) - Auteur collaborateur de Voicebot.ai - Responsable produit chez Amazon	8 juillet 2021	251
10.	<u>Stanley Goffinet</u> <u>Fonction :</u> - Étudiant en master 2 en sciences informatiques à l'école polytechnique de Louvain - Mémoire publié: "Controlling security rules using natural dialogue: An application to smart home care"	9 juillet 2021	264

Cas d'usage n°1 : Aide à la conduite

1. Interview de Oxana Gouliaéva

Bonjour. Tout d'abord merci beaucoup d'avoir accepté de participer à cet entretien. On va peut-être commencer par se présenter brièvement. Donc moi c'est Pauline et Alice est ma co-mémorante. Nous sommes en train de co-rédiger un mémoire au sujet des assistants vocaux. On est étudiantes en dernière année de master et on étudie le management à Louvain-la-Neuve. Notre mémoire a pour objectif de définir la valeur ajoutée de l'assistance vocale dans 3 cas d'usage concrets. C'est l'aide à la conduite, l'apprentissage d'une langue et les services de secours à distance. Donc on vous a contactée pour mieux comprendre tout ce que représente l'intelligence artificielle, les assistants vocaux et aussi pour avoir votre point de vue sur le premier cas qu'on étudie, qui est l'assistant vocal dont les voitures sont équipées. Voilà, je vais vous laisser la parole pour vous présenter, votre parcours et votre lien avec les assistants vocaux.

Oui bien sûr, avec plaisir. Alors, ma dernière expérience n'a probablement rien à voir avec les assistants vocaux. J'ai fait un peu de recherche autour de l'intelligence artificielle et des *use cases* qui sont liés à mon poste précédent chez Accenture Avanade. J'ai également fait un master exécutif à Sciences Po avec comme sujet les assistants vocaux et comment ils changent le marketing des marques, qui dans une version modifiée et enrichie est devenu le livre, chez Dunod, "La révolution des assistants vocaux". Je suis également membre et un petit peu active dans quelques associations ou réseaux qui tournent autour des technologies vocales dont les assistants vocaux font partie.

D'accord merci, donc là je pense que l'on peut rentrer dans le vif du sujet avec quelques questions que l'on a préparées. Alors selon vous, quelles sont les avancées actuelles des voitures dotées d'une assistance vocale ?

Question piège. Je ne sais pas, j'avoue que dans les derniers 6 mois j'étais trop prise sur d'autres sujets pour suivre les dernières évolutions en sachant que je ne pense pas que cela soit encore vachement avancé. Effectivement, les constructeurs automobiles, depuis un moment déjà, et pour de bonnes raisons, cherchent à intégrer l'assistance vocale dans leur cockpit, c'est-à-dire

en natif. Ce n'est pas sur le smartphone ou un autre *device* que l'utilisateur pourra amener avec lui, mais en natif pour proposer une expérience de conduite. Peut-être même une expérience tout court, notamment dans des régions, des pays ou d'autres circonstances de vie où les gens restent beaucoup de temps dans les voitures. Alors quelles sont les dernières avancées, j'aurais du mal à le dire mais je pense que j'ai un bon indice. Le fait que chez Nuance⁸⁴, l'année dernière je crois, la partie dédiée en *automated market*, est devenue une entité à part entière. C'est-à-dire que l'on voit une nette marge de progression sur ce marché là pour créer des solutions dédiées uniquement à ce marché-là. Donc je pense que ça c'est plutôt un signe fort sur un acteur historique des technologies vocales liées au marché, et le marché c'est une variable très importante pour tout ce qui est l'innovation. Je pense qu'il y a effectivement un gain quelque part à court terme ou moyen terme sur ce marché-là. Je ne veux pas me prononcer sur la technologie parce que finalement elle est la même, il s'agit uniquement de la capacité de telle ou telle marque automobile ou du constructeur automobile de s'emparer d'une technologie. On peut avoir le même discours sur les voitures autonomes.

Est-ce que vous pensez que l'assistance vocale pourrait apporter une plus-value justement aux voitures qui en sont équipées ?

Elle le fait déjà, comme tous les use cases qui sont peut-être les plus adoptés, on se moque très souvent de « Ok Google, c'est quoi la météo ? », mais en réalité c'est un gain de temps et on se rend compte que l'assistance vocale dans toutes les circonstances finalement répond à certaines contraintes, on va dire celle du gain de temps ou de mains occupées. Et dans la voiture on a quelque part les deux d'une certaine manière, on a bien sûr les mains occupées mais on a surtout envie d'interagir avec d'autres personnes, et d'autres restrictions qui peuvent intervenir. Donc il y a clairement une valeur ajoutée.

Ensuite, je ferais quand même une petite nuance entre valeur réelle et valeur perçue, tout est dans la perception. C'est un peu comme le prix et comme la valeur, mais en fait c'est la perception. Donc ça c'est un point important. On parle de ce cas-là dans le livre avec ces questions : Est-ce que les clients sont prêts à payer pour ce surplus de valeur qui leur est proposé ? Est-ce qu'ils en font un critère de choix ? Est-ce qu'on va choisir telle ou telle voiture parce

⁸⁴ Nuance est une multinationale américaine spécialisée dans les logiciels informatiques, et qui fournit des services de reconnaissance vocale et d'intelligence artificielle.

que le système audio est plus performant ou parce que leur assistant vocal est intégré, est natif, et est celui de votre constructeur et pas votre Google ou Alexa ou autre ? Donc il y a assez de choses à avoir en tête quand on pose ce genre de question à mon sens.

Ok merci, et justement lié avec cette question, quelle serait la plus grosse valeur ajoutée, toujours avec cette question de perception ? Pour nous, la première chose qui nous est venue c'est la sécurité évidemment, une sécurité supplémentaire dans la conduite. Est-ce que vous le percevez de cette façon-là aussi ?

C'est-à-dire vous parlez surtout quelque part de l'avantage qui est offert par l'interaction vocale. Par là je dirais que ça dépend encore une fois du cas d'usage. Ça dépend parce qu'on peut demander plein de choses à son assistant, on peut bien sûr demander de freiner, de baisser sur la vitesse, vous voyez à terme avec une voiture autonome on peut réagir comme ça. Pour l'instant s'il ne fait que remplacer le smartphone, ou utiliser le GPS, probablement la première *feature* pour laquelle la voix s'applique, oui certes. Mais encore une fois cela va dépendre de plusieurs choses, ça va dépendre de la performance de l'assistant en général aussi bien en compréhension de ce qu'il dit, de la pertinence de la réponse ou de l'action, si c'est une action qui est demandée. Et encore une fois du cas d'usage, s'il s'agit ensuite de discuter avec quelqu'un juste composer le numéro de manière vocale, oui on va gagner du temps effectivement, on ne va pas lâcher le volant éventuellement, mais on ne lâche plus, on a des écouteurs et même des décodeurs sans fils.

Votre question concerne d'un côté l'interface et l'autre côté l'usage qui est fait ensuite de la voix. On discute dans la voiture et ça nous distrait quand même, donc sécurité je ne suis pas sûre, encore une fois ça dépend de ce que vous faites avec la voix. Je pense que pour moi, ça rejoint la partie expérience au sens très large, pour moi ça veut dire que notre perception de tout ce que nous consommons en tant qu'utilisateur est donc plus *drivé* par la partie expérience qu'on retient, ce n'est plus la qualité de la conduite si on choisit entre une Renault, une Peugeot et même peut-être une Audi, c'est moins la qualité de la conduite, en tout cas pas pour tout le monde, mais c'est autre chose. C'est, par exemple, le confort des sièges, ou autre chose. L'interaction vocale c'est une interface, on apprend ce que la nouvelle interface peut nous donner avec toutes nos contraintes. On parle très souvent des gens on va dire normaux, mais on ne se rend pas compte qu'il y a des gens avec des handicaps peut-être plus différents, qui peuvent

être visuels, qui peuvent concerner l'audition, qui peuvent être des handicaps liés à la motricité. Il y a plein d'autres choses, il y a l'âge également qui est concerné.

Donc vous voyez dans la voiture, la sécurité je ne suis pas si sûre que ce soit le premier truc. Il faut voir ça comme quelque chose de plus large, on n'a plus la perception et la même expérience qu'on fait du monde qui nous entoure et des expériences, des produits, et de tout ce que nous consommons si vous voulez. Et cette interaction vocale, à mon avis, simplifie dans beaucoup de cas la technologie et l'accès à la technologie. On n'est pas obligé de taper à l'ordinateur très très vite, si vous ne savez pas taper avec les dix doigts à l'aveugle, alors vous les dictez. Si la technologie est suffisamment performante, il y aura moins d'erreurs. Et finalement c'est ça qui vous fait avancer, vous gagnez du temps, on est toujours dans un gain de temps. Pour moi c'est plutôt le gain de temps en général qui est le premier avantage, le premier bénéfice des technologies vocales, des assistants qui font partie des technologies vocales.

Selon vous, vous voyez plus l'assistant vocal dans les voitures comme un gadget qu'une innovation qui va devenir indispensable dans les voitures ?

Non, je n'ai pas dit que c'était un gadget, en revanche, c'est une autre expérience qu'on vit de la voiture. Je n'ai pas dit que c'était un gadget, c'est une interface qui fait que vous ne touchez pas l'écran, vous ne touchez pas à un bouton, vous ne tournez pas un truc. Éventuellement, vous ne vous écartez pas du volant etc. Il y a des éléments de sécurité, mais ce n'est pas un gadget, encore une fois ça dépend de ce que vous faites en utilisant des interfaces invisibles. Finalement comme ce n'est qu'une interface, d'une certaine manière il faut voir quelle est l'action que vous demandez à l'assistant d'exécuter ou des informations à fournir. Donc c'est ça qui va être intéressant, si vous pouvez le faire, si vous pouvez accéder à plus de services différents qui concernent la conduite de la voiture mais également des choses complètement externes, comme anticiper votre arrivée à la maison pour lancer le chauffage, commander une pizza ou autre, c'est finalement ça et ça ne concerne pas que la voiture. Mais tout simplement j'ai envie de mettre entre parenthèses un peu la voiture parce que je ne sais pas si la sécurité est le premier avantage dans la voiture. J'ai du mal à me projeter uniquement dans ce cas-là. Non ce n'est pas un gadget, je ne crois pas.

OK. Je reviens sur ce que vous disiez sur le fait que l'assistant vocal simplifie la technologie. J'ai une question par rapport à ça, pour vous où se situent les consommateurs en termes de l'adoption de l'assistant vocal ?

C'est très variable selon la géographie, l'âge et le développement de la technologie dans la langue de l'utilisateur. Ça je vous renvoie plutôt aux études qui sont faites, on peut les piocher un peu à droite à gauche, ça existe. Je ne suis pas forcément à jour mais je pense qu'il y a eu une espèce de certains plateaux depuis peut-être un an, où l'adoption a cessé peut-être de progresser aussi vite. En tout cas aux États-Unis même où le quart a été franchi pour les 30% d'adoption, mais probablement lié à la maturité de la technologie. Apparemment ça commence à stagner un petit peu à ce niveau-là. En revanche ce qui est intéressant, mais là c'est pas nouveau non plus, c'est quand même que le décollage de ces technologies, même si je suis pas tout à fait en confiance avec les sources détaillées, en Chine où il y a ses propres acteurs du vocal, des technologies propres qui s'adaptent à tous les dialectes de Mandarin et de Cantonais et autre. Avec l'usage du mobile, qui est le premier terminal que tout le monde a dans sa poche, tout le monde n'a pas un ordinateur mais tout le monde a un portable, un smartphone, là l'usage vocal est répandu également. Apparemment ça fonctionne très bien, je ne vais pas vous donner forcément des chiffres mais ça dépend encore une fois de la maturité de sa population à tout type de nouveaux usages très souvent liés à l'âge, mais pas que. Donc la présence des acteurs nationaux simplifie d'ailleurs la partie technologique, en adoptant le langage. Alexa est moins bonne en français qu'en anglais, ce qui fait que le niveau de la technologie étant meilleur, les gens adoptent plus facilement.

Une autre question de nouveau un peu plus tournée vers la voiture, selon vous quels seraient les freins les plus importants à l'adoption de l'assistant vocal dans la voiture ?

Je pense que ma réponse sera encore une fois générique, les mêmes freins que peut-être ailleurs, dans d'autres circonstances. Est-ce qu'on fait la différence entre n'importe quel assistant que j'ai dans ma poche ou celui du constructeur automobile ? Certains vont préférer l'un ou l'autre selon d'un côté la perception de la technologie elle-même, et de l'autre côté selon le côté intuitif, usabilité de cet assistant. Et là on est dans les questions de *design*, peut-être que celui de la voiture sera mieux *designé*. Finalement on ne comprend pas que c'est une technologie, c'est une voix qui nous répond. Ça nous apporte un certain flou, on comprend que c'est une machine

mais en même temps ? s'il a une voix qui n'est pas mécanique, c'est assez perturbant psychologiquement. Si cette expérience est intéressante et proche de l'humaine, on va peut-être bien l'aimer. Vous savez que Siri, qui semblait un peu bête au début, c'était volontaire. Il était un peu bête parce qu'il fallait couvrir certains défauts de la technologie. Quand il ne savait pas répondre concrètement et précisément à une question, on utilisait toutes sortes de phrases qui tournaient autour du pot mais qui le rendait humain. Finalement, le fait qu'il ne connaissait pas, ne nous dérangeait pas tant que ça. Donc il y a cet élément extrêmement important dans l'expérience vocale, on a des critères de choix ou de préférence pour tel ou tel assistant. Encore une fois, que ce soit dans la voiture ou non, je pense que le fait de gagner du temps et d'avoir les mains occupées donc ne pas forcément toujours vouloir s'arracher du volant, est une contrainte importante. Mais j'aurais du mal à vous donner une réponse plus précise que ça.

Pas de soucis. Et au niveau des utilisateurs, est-ce-que la gestion de leurs données fait partie des freins à l'utilisation de l'assistant vocal ?

Je pense que ça en fait partie, encore une fois je vous renvoie à ce que j'ai dit tout à l'heure. Pour l'adoption, ça dépend de la région du monde, ça dépend du niveau de *digital nativity*, souvent lié à l'âge mais pas que. Ça dépend beaucoup de régions du monde, de la présence ou pas de l'utilisation ou pas d'acteurs nationaux donc avec la technologie nationale, et de l'attitude face à la technologie en général. C'est en Europe, et surtout en France, que les freins sont les plus importants. Ailleurs en Europe, en fonction des pays, ce ne sont pas du tout les mêmes chiffres, c'est pas du tout homogène. Aux États-Unis, ces freins sont beaucoup plus *light*. Et en Russie, le taux d'adoption de l'assistant on va dire "local", n'est pas super élevé parce qu'il n'est pas super performant. Encore une fois, en Chine, apparemment, beaucoup de choses fonctionnent, en revanche on utilise plutôt la partie commande vocale que l'assistant en tant que tel, c'est juste pour dire "ouvre la porte" mais ça suffit. Ça permet de gagner du temps, de faire plus de choses en même temps. Les gens ne s'en préoccupent pas tellement parce que la question ne se pose pas ou parce qu'ils ont à la limite plus confiance parce qu'ils utilisent leur assistant national qui appartient à l'État. En Chine, la question ne se pose pas ou en tout cas beaucoup moins, pas avec la même gravité et pas pour les mêmes raisons qu'en France. Je suis sûre que si Alexa ou Google était une invention française, on en aurait été à 90%. Pardon, vous êtes en Belgique et je vous donne des exemples français, parce que la France est effectivement celle qui résiste le plus à la partie des données.

Une question plus orientée maintenant vers les GAFA. Est-ce que vous pensez que si justement on intègre des assistants vocaux dans les voitures, est-ce qu'il s'agira d'insérer les GAFA ou ce serait plus les assistants spécifiques des constructeurs de voitures ?

Les constructeurs de voitures ont tout intérêt à développer leurs assistants. Ils le font même si rien ne les empêche d'expérimenter avec d'autres assistantes du marché, ce qu'ils font aussi. Le même constructeur peut avoir deux ou peut-être même trois projets en parallèle. Les GAFA, étant juste les plus performants parce qu'ils ont le cloud, ça leur permet d'affiner les *data*, de perfectionner l'entraînement des algorithmes etc. Techniquement ils sont naturellement plus forts que les autres. Surtout dans la partie qui concerne le B2C, le grand marché, on va dire la consommation large. Il y a des usages également B2B avec des acteurs B2B dont on peut parler également entre parenthèses. D'autres acteurs néanmoins existent, il y a toutes les briques de l'intelligence artificielle dont les technologies vocales, les reconnaissances vocales, traitement de langage naturel, text-to-speech etc, qui sont développées. Il y a des solutions alternatives qui sont développées par des start-ups ou des entreprises plus matures. Je pense notamment à un service avec Nuance et Sonos qui se développe également.

Il y a plusieurs sujets qui font que c'est pas forcément le plus fort qui va gagner, c'est pas forcément le premier. On voit déjà qu'entre les pays, les régions, Google et Alexa n'ont pas la même force. Il y a des raisons commerciales et stratégiques. Il y a aussi plein d'autres raisons derrière. Encore une fois les constructeurs automobiles ont toutes les raisons de développer leur propre technologie pour conserver cette relation directe qu'ils ont avec l'utilisateur, qui est le conducteur, pour éventuellement étendre cette expérience au-delà de leur enjeu et au-delà du vocal, et de garder la relation comme toutes les entreprises, et ne pas être intermédiés quelque part par les GAFA. Mais effectivement les GAFA sont bien placés pour, pour être effectivement en tête de liste. Ou avec les actions antitrust etc, ils peuvent très bien dans quelques années être un peu séparés et même si on parle de Google en fait c'est Alphabet qui aura son entité, qui va s'occuper de cette partie par exemple des voitures autonomes, du vocal etc, et toutes ces choses-là à priori peuvent également permettre de contourner les parties monopoles et antitrust.

Donc oui, mais en même temps rien n'empêche les acteurs automobiles de contourner ça. Leurs assistants pourront peut-être commencer par une bonne expérience en automobile et selon l'essor que la technologie prendra et la vitesse de l'adoption et des cas d'usage qui vont se

développer, leur assistant peut rester complètement dans la course. Encore une fois il y a des questions d'interopérabilité dont on parle de plus en plus. Rien ne nous empêchera peut-être dans quelques années, de la même manière, d'interagir avec plusieurs assistants.

J'ai peut-être une question, parce que on a entendu parler que Waze, enfin l'algorithme de Waze, avait été un peu tronqué par la publicité. C'est-à-dire qu'il ne prenait pas le chemin le plus opportun pour le conducteur mais par exemple choisissait un chemin qui l'orientait vers une publicité ou un magasin. Donc est-ce que ce genre de choses arrive aussi dans les assistants vocaux ?

J'ai l'impression que vaguement j'ai vu quelque chose de ce genre passer mais franchement je n'ai pas creusé plus que ça, je ne connais pas cette histoire avec Waze. Encore une fois, on n'est pas dans le vocal là, on est dans un modèle économique, oui effectivement ça peut le faire de la même manière que Google vous sort d'abord tous les *snippets* qui sont payés ou pas payés au début, qui ne vous renvoie même pas les médias directs et qui vous donne des réponses tout à fait pertinentes en bas de page, de plus en plus en bas de page. On est plus dans le modèle économique, et encore une fois pourquoi on va utiliser tel ou tel assistant.

Est-ce que le fait qu'il ne sert pas mes intérêts à moi mais celui qui a payé la pub, va être un critère de choix dans l'adoption de tel ou tel assistant ? Vous pouvez refuser d'utiliser Waze en disant non je vais chercher un autre algorithme. Je vais faire confiance peut-être d'ailleurs à mon assistant vocal de conduite dans la voiture. Vous voyez ça peut devenir un critère de choix ou de non-choix. Les gens qui sont sur Apple, ils vont se débrouiller autrement. Il y a également les questions de l'OS⁸⁵, même au-delà d'un assistant en particulier. Si on est sensible par exemple aux données ou au modèle, on peut faire des choix qui intègrent ces critères-là qui sont importants pour nous. Donc pour moi c'est ça rien à voir avec le vocal, c'est à voir avec n'importe quel moteur de recherche.

Peut-être une autre question, quelles sont les demandes qui restent trop complexes à traiter pour un assistant vocal ? Par exemple, on avait lu que des demandes qui cumulent plusieurs paramètres ne sont pas encore vraiment très performantes. Par exemple demander “trouvez-

⁸⁵ Operating System, ou système d'exploitation en français, est un logiciel qui, dans un appareil électronique, pilote les dispositifs matériels et reçoit des instructions de l'utilisateur ou d'autres logiciels (ou applications). Dans un ordinateur, le système d'exploitation gère le ou les processeurs ainsi que la mémoire (Futura Tech, s.d.).

moi un fast-food qui n'est ni McDonald ni Burger King". Est-ce que cela va s'améliorer pour pouvoir composer plusieurs paramètres en même temps ou est-ce que c'est encore trop complexe pour le moment pour un assistant vocal ?

Encore une fois vous parlez des technologies qui sont complètement différentes, c'est-à-dire qu'il y a plusieurs technologies qui sont proposées. Il y a des technologies qui sont plus ou moins avancées là-dessus, qui visent cette complexité de la demande spécifiquement et d'autres qui ne le font pas. Notamment, je pense à SoundHound, qui a fait sa réputation entre autres sur le fait qu'ils peuvent traiter des demandes très complexes avec plusieurs critères à l'entrée, typiquement "trouve-moi un fast-food ni McDo ni Burger King, qui est ouvert maintenant mais qui a un menu avec ceci, et prêt dans 10 minutes". Il peut intégrer plusieurs choses, s'il est lié à leur technologie maison, propriétaire interne, qui entre autres, tout simplement commencent à traiter la demande avant que la phrase soit finie. D'autres attendent que la phrase soit finie pour avoir l'ensemble des entités et des intentions qui vont figurer dans la requête pour lancer le traitement. Entre autres, ils commencent à la traiter dès le début.

Donc en fait ça dépend vraiment des technologies, de la même manière que des burgers sont différents d'une maison à l'autre, que ce soit fast-food ou pas, ceux que vous pouvez trouver dans une brasserie et peut-être même des restaurants. Ce n'est pas exactement le même mais pourtant on va l'appeler de la même manière. On va peut-être écouter les assistants vocaux, il y aura autant de nuances, et en tout cas plus de nuances que de freins que nous connaissons au grand public.

OK, merci, je pense qu'on a plus ou moins fait le tour de toutes les questions que l'on voulait vous poser. Est-ce-que vous de votre côté vous avez peut-être quelque chose à ajouter, une piste sur laquelle nous orienter ?

Je pense que si vous tournez autour des usages notamment liés à la voiture, encore une fois ça dépend de la technicité que vous recherchez etc. donc non je pense que vous vous posez les bonnes questions.

Merci beaucoup, en tout cas votre livre nous a déjà pas mal aidées pour notre revue de littérature.

Écoutez c'est très bien, je suis contente, et si vous avez d'autres questions qui viennent ou vous trouvez que je peux vous aider, n'hésitez pas.

Merci beaucoup, et encore merci pour votre temps.

2. Interview de Ariane Nabeth-Halber

Bonjour, tout d'abord merci d'avoir accepté de participer à cet entretien. On va peut-être se présenter brièvement avant de rentrer dans le vif du sujet. Donc je m'appelle Pauline, c'est avec moi que vous avez été en contact et Alice avec qui j'écris mon mémoire. Notre mémoire a pour objectif de définir la valeur ajoutée de l'assistance vocale dans trois cas d'usage que l'on a choisis et qui sont : l'aide à la conduite, la traduction et les services de secours à distance.

On vous a donc contactée pour mieux comprendre tout ce que représente l'intelligence artificielle, les assistants vocaux mais aussi pour avoir votre point de vue sur ce premier cas qu'on étudie donc on aimerait parler en particulier des assistants vocaux dont les voitures sont équipées. Voilà, je peux vous laisser la parole pour vous présenter, parler de votre parcours et peut-être nous parler du lien que vous avez avec les assistants vocaux.

De mon côté, Ariane Nabeth-Halber. Moi je travaille dans le monde de la parole depuis les années 90. J'ai d'abord travaillé dans la recherche en reconnaissance de parole dans des labos comme Ater au Japon ou Thales en France et le labo de Télécom Paris, où j'ai fait ma thèse. Donc j'ai travaillé sur la reconnaissance de parole, la section de mots-clés, la compréhension de parole et l'analyse du langage. En fait, les deux sujets sont assez connexes même si ce sont des communautés assez différentes. Enfin jusqu'à il n'y a pas si longtemps, les communautés n'étaient pas différentes mais ça commence à converger un peu.

Ensuite, je suis passée dans l'industrie. Depuis les années 2000, je suis dans l'industrie des serveurs vocaux, des assistants vocaux, de la transcription de parole, etc. Je suis passée notamment chez Nuance, qui a récemment été rachetée par Microsoft, qui est dans le moteur. Il a longtemps été derrière Siri. Puis, pour d'autres sociétés, notamment en France.

Aujourd'hui je suis responsable de la ligne de produits *speech*, donc transcription de parole et reconnaissance de parole dans une société française qui s'appelle Bertin IT. J'ai choisi de passer maintenant à un autre défi, dans une autre société française. Elle s'appelle ViaDialog et elle est dans le monde de la relation client. Elle a aussi sa capacité de R&D. Donc ils travaillent beaucoup sur les assistants vocaux avec des recherches spécifiques sur la synthèse vocale,

l'émotion, la reconnaissance de parole, etc. Ils ont une vision de bout en bout du métier et des utilisateurs, que ce soit l'utilisateur téléconseillé dans le centre de contact ou l'utilisateur appelant.

Voilà pour ma présentation. Après qu'est-ce que je peux vous dire dans mes différents rôles au sein de l'industrie ? Je pense que j'ai toujours joué un petit peu les ponts entre le monde de la recherche et le monde des applications. C'est ça qui m'a passionnée et qui a fait que je ne suis pas revenue tout de suite à la recherche. Au début j'avais dit à mon jury “ne vous en faites pas, je vais faire 6 mois dans l'industrie et puis je reviens vite”. Je ne suis pas revenue et je n'en ai pas envie. Mais je ne suis pas vraiment perdue pour la recherche. Le fait est que mon rôle a un petit peu évolué. Au lieu d'être vraiment chercheuse dans un labo, je suis plutôt ce pont entre des équipes de recherche et des besoins clients. En fait c'est ça que je trouve passionnant et c'est assez important : d'avoir la compréhension des deux pour comprendre comment on peut faire converger et comment on peut arriver à des nouveaux usages. Dans mes recherches chez Bertin IT, où j'ai passé 7 ans, j'ai eu la chance de pouvoir déployer des services qui étaient très innovants et où on explorait finalement aussi bien côté IT que le côté usage.

Ok merci, je pense que l'on peut commencer avec les questions du guide d'entretien. On va commencer par la première question : que pensez-vous des avancées dans les assistants vocaux qui se trouvent dans les voitures ? Enfin où en sommes-nous dans les avancées actuelles ?

OK, alors c'est une bonne question. Moi j'ai vu les premiers assistants vocaux en voiture quand j'étais chez Nuance et on commençait à développer ça en 2001-2002 avec des beaux contrats, qui avaient été signés avec Ford et Mercedes. C'était les choses qui étaient réservées plutôt aux modèles haut de gamme des voitures de ces constructeurs-là. C'est pour une raison toute bête, qui est la question du bruit dans l'habitacle. C'est-à-dire que seuls les modèles haut de gamme avaient suffisamment d'insonorisation pour permettre à la techno de fonctionner.

Ensuite, j'ai été confrontée à nouveau à ce monde-là vers 2013 ou 2014, pour deux raisons. D'une part parce que le moteur médiatique avait intéressé certains constructeurs français, qui voulaient voir si ça pouvait servir de base pour faire des assistants. Donc, avec mon équipe, j'avais planché un peu sur le sujet pour voir si on pouvait répondre à leurs demandes.

En même temps, à ce moment-là, je regardais aussi comment évoluait la société française, qui s'appelle Kalisma, qui était une spin-off des laboratoires d'Orange, de Télécom. Ils avaient évolué et commençaient à être assez bien positionnés dans le monde automobile. Ils se sont d'ailleurs fait racheter par VoiceBox, une société qui, à l'époque, était concurrente de Nuance sur l'assistant automobile. Elle s'est d'ailleurs fait racheter par Nuance.

À ce moment-là, j'avais un petit peu regardé, de façon très ouverte, avec les gens de Kalisma et MediaSpeech, ce qu'on pouvait faire. En réalité, (je vous donne le point de vue du monde technique technologique parce qu'à ce moment-là j'avais pas forcément le point de vue côté acheteur ou le côté client mais en tout cas pour les fournisseurs de techno) je m'étais rendue compte que la barrière à l'entrée n'était pas tant la technologie de reconnaissance vocale. La barrière à l'entrée c'était toutes les règles culturelles, dialogiques, métiers dans toutes les langues. Parce que aussi bien TechSpeech que MediaSpeech fonctionnaient dans beaucoup de langues (enfin tout est relatif, disons dans une dizaine/quinzaine de langues). Il fallait donc rajouter 4 ou 5 langues mais ce n'était pas forcément ça l'effort principal. Ce n'était pas forcément ça qui était le plus coûteux. Le plus coûteux c'était de développer tous les dialogues en contexte, dans toutes les langues. Quand on demande un restaurant, une station-service dans un pays, il y a toutes sortes de conventions qui font que quand on demande un restaurant, il faut avoir, derrière, un dialogue qui va répondre "plutôt un fast-food, plutôt un italien ?". Il y a des points communs entre les pays et puis il y a des choses qui sont spécifiques. Donc, mettre au point tous ces petits dialogues en fonction du contexte, en fonction de la langue, du pays, de la culture du pays, c'est ça qui avait été la barrière à l'entrée la plus grande.

Nuance et VoiceBox possédaient déjà ça, ils avaient déjà travaillé là-dessus. Ça rendait la tâche difficile pour les nouveaux entrants technologiques, même ceux qui avaient les bonnes technologies de reconnaissance ou de synthèse. C'était ce côté localisation culturelle, qui était l'effort initial à fournir et qui faisait barrière à l'entrée dans le secteur.

Voilà, je vous ai dit les points où j'ai rencontré de près le domaine des assistants vocaux en automobile. Je ne les ai pas rencontrés de très près récemment. Ce qui s'est beaucoup popularisé, c'est tout ce qui est assistant vocal type Alexa, Siri et Google, etc. Ce sont les assistants vocaux sur les téléphones mobiles et les enceintes connectées. Ces enceintes connectées, évidemment, font beaucoup de bruit, sont très visibles. Elles ont quand même gagné des millions

d'utilisateurs en peu de temps. Alors que la question de l'utilité était quand même un point d'interrogation au départ. Donc c'est surtout ça qui a attiré l'attention et sur lequel il y a beaucoup d'informations.

J'ai été confrontée de façon plus précise d'un point de vue technologique, à des demandes provenant d'institutions qui ne souhaitent pas avoir recours à des services connectés. Il y a des services non souverains qui demandent à avoir leur propre assistant vocal pour des ministères ou des services souverains. Moi j'ai été confrontée à ça mais dans le cas de l'automobile, non. Je vous dirais qu'après mon expérience dans les années 2013-2014, le fait que la barrière à l'entrée n'était pas la technologie, c'était le fait de développer tous ces dialogues associés à tous ces pays, toutes ces cultures, je n'ai pas encore vu le côté offre évoluer. J'ai bien envie de vous renvoyer la question : est-ce qu'aujourd'hui vous voyez Alexa, Google home etc, déployer leur pendant pour le monde automobile, c'est une vraie question.

C'était justement un peu le sujet d'une de nos questions : est-ce que ce sont vraiment les GAFAs qui vont être intégrés dans les voitures ? Ou est-ce que les constructeurs vont construire des voitures équipées de leur propre assistant vocal ?

J'ai un début de réponse, je réfléchis à ce que je connais vraiment et ce que j'ai pu voir récemment. J'ai quand même été sollicitée par des sociétés comme Faurecia⁸⁶. C'est une grande société qui est un équipementier pour les constructeurs automobiles et qui développe beaucoup d'électronique et de services embarqués pour les constructeurs. Ils ont évidemment l'innovation, ils regardent toujours ce qui se fait. Bien sûr, ils se sont intéressés au vocal. On a eu quelques discussions mais pas très poussées sachant que moi à ce moment-là, dans mon rôle, c'était pas un secteur stratégique. En tout cas, je n'ai décidé d'investir du temps, de mettre des efforts là-dedans. Indépendamment même de la techno, il y a eu quelques discussions.

Faurecia, par ailleurs, expérimente aussi des choses aujourd'hui avec des assistants type Alexa, etc. Le principe qui est utilisé, si j'ai bien compris, c'est que l'assistant vocal développé par Faurecia, va être un assistant spécifique sur les fonctions que connaît bien cet équipementier (tout ce qui va gérer l'électronique de bord, par exemple, demander à déclencher les essuie-glaces, demander à moduler la clim ou le chauffage, etc.). C'est de l'électronique de bord, que

⁸⁶ Faurecia est un fournisseur automobile mondial français

maîtrise Faurecia. Donc eux ils développeraient, ou ils ont développé, une sorte d'assistant vocal spécialisé sur ces fonctions liées à l'autonomie.

Par ailleurs, il se connecte avec un Alexa ou autre pour tout ce qui est fonction plus générique. Il y a une espèce de complémentarité entre ces deux assistants vocaux, qui se passent la main. Ils auraient pu choisir de développer des *skills* Alexa mais je ne crois pas que c'est ce qu'ils ont fait. Ils ont choisi plutôt cette métaphore des deux assistants vocaux qui se passent le relais. C'est un grand débat aujourd'hui dans les assistants vocaux : quelle est la bonne métaphore et qu'est-ce qu'on fait quand il y a plusieurs *skills*, plusieurs assistants vocaux ou plusieurs *branding*? Est-ce que c'est bon ou pas bon pour l'expérience utilisateur ?

Je sais qu'il y a des avis partagés là-dessus. Certains ergonomes comme Ahmed Bouzid en ont parlé. C'est quelqu'un de très intéressant, qui a travaillé notamment chez Amazon et qui a maintenant sa société Witlingo. Il est très intéressant, il anime plein de débats, de conférences. C'est une super source d'information. Lui, son point de vue là-dessus c'est que c'est contraire à la bonne expérience utilisateur, cette idée de faire ressortir ce passage de relais. Il n'est pas pour le fait de dire « je vais passer maintenant la main à ma collègue Faurecia » ou « j'ai passé la main à ma collègue Alexa », etc. C'est toutes ces questions qui font ressortir le *branding* des marques des assistants vocaux. Il comprend cette notion de visibilité, de publicité, d'image et de marketing derrière la marque. Cependant, sa conviction, c'est que d'un point de vue de l'expérience utilisateur, la fluidité de l'interaction, l'intérêt même de l'utilisabilité de ces interactions vocales, ça n'apporte rien. Ça serait plutôt une gêne dans la fluidité de l'interaction.

Au fond, l'interaction vocale est là pour rendre certains services à l'utilisateur et il n'a pas forcément besoin de savoir par quelle marque ça lui est apporté. C'est un petit peu comme quand on vous diffuse des pubs des services et que ça vous fait perdre un petit peu votre temps. C'est son point de vue, après je pense que c'est sujet à débat. En tout cas, la dernière fois que j'ai été confrontée de près à cette question de l'assistant vocal en voiture, c'était au sujet de ces travaux mais qui sont encore l'innovation chez Faurecia. Je ne pense pas que ce soit quelque chose de complètement figé. Je crois qu'ils continuent à travailler sur le sujet. Ils avaient des personnes qui intervenaient lors de la dernière conférence Voice Tech Paris. Je pense qu'il y avait une personne qui s'appelle Soledad Bedu qui était intervenue. Elle parlait spécifiquement sur ce sujet-là, l'ergonomie de l'assistant vocal en voiture. C'est une personne que je connais bien

parce qu'elle avait été une collaboratrice de Bertin IT à un moment donné. J'étais heureuse de la revoir intervenir et si vous avez besoin, je peux vous mettre en contact avec elle, il n'y a pas de souci. Elle pourra peut-être vous parler aussi et là pour le coup elle est vraiment sur ce sujet-là : la conception d'assistants vocaux dans les voitures.

Ah oui merci, ça serait super intéressant en tout cas.

Oui, alors j'ai pas bien répondu à votre question mais voilà. La conclusion c'est probablement que oui, il y a une porosité entre les assistants vocaux qui sont développés qui sont devenus très visibles aujourd'hui dans le paysage des biens de consommation électroniques. Il y a une porosité entre ça et les assistants vocaux automobiles. Est-ce que ça va changer complètement la donne par rapport à la domination quasi hégémonique de Nuance, Voice Box pendant pas mal d'années ? C'est encore tôt pour le dire. J'aurais tendance à dire que dans ces marchés là, ce qui compte le plus ce n'est pas la techno, ce n'est pas non plus forcément cet effort dont je vous parlais sur les dialogues. C'est plutôt la prise en compte des données culturelles pendant les dialogues qui doivent être livrés dans tant de langues, tant de pays en même temps. Ça reste bien sûr un sujet mais très souvent ce qui va faire aussi l'adoption, le développement de ce type de techno, c'est le business model tout simplement. Pourquoi les enceintes connectées s'en sortent comme ça et ont un tel succès ? Parce qu'elles sont portées par des sociétés comme Google ou Amazon, Apple, Samsung. Pourquoi ces gens-là ? Tout simplement c'est leur business model.

Les enceintes connectées, c'est une nouvelle façon d'ouvrir le marché pour concurrencer l'hégémonie de Google sur la recherche web. C'est-à-dire que vous avez un acteur qui a une hégémonie quasi mondiale. Je crois qu'au niveau mondial, à part en Chine, 90% des requêtes sur le web sont traitées par un Google. Peut-être 96% aux États Unis parce qu'il y a Microsoft qui essaye toujours de concurrencer et en Europe, 99%. Enfin, il n'y a quasiment aucun concurrent. Il y a Qwant mais bon c'est quelques valeureux qui, politiquement, ont choisi ce moteur. C'est une action militante. Tout d'un coup, vous avez les enceintes connectées et ça a créé un nouveau marché où il n'y a pas d'hégémonie de faite. Bien sûr, Google ne veut absolument pas perdre sa domination sur ce marché. Il veut connaître les désirs et demandes des gens et monétiser ça. Cependant, on voit un géant comme Amazon qui s'y met aussi. C'est très logique parce qu'au fond Amazon a un *business model* très compatible avec le fait d'être en

première ligne sur les demandes et les désirs des gens. Son modèle c'est aussi de leur livrer les objets, les biens, etc. qu'ils voudraient commander. Puis Amazon s'est mis dans le contenu aussi avec Amazon Prime. C'est le *business model* d'Amazon donc il y a une logique complète dans le fait que Amazon se mette sur ce nouveau terrain de jeu.

Samsung est là en tant que fournisseur de *devices*, de *smartphones*. Il a quand même cette capacité à avoir cette base d'utilisateurs et à essayer aussi d'être en première ligne. Ils ont leur propre assistant vocal comme Apple avec Siri. On voit bien que les gens qui se mettent sur ce marché là, sont ceux qui ont le *business model* qui va bien avec. Alors que d'autres essaient de s'y mettre. On a vu la malheureuse tentative d'Orange avec Djingo, qui malheureusement n'a pas tellement fonctionné. Ils ont soumissionné il n'y a pas très longtemps. C'était leur propre enceinte connectée avec un appareil en espèce d'entente avec Amazon ou Alexa ou un autre. Ils avaient un accord avec un assistant vocal existant américain. Puis, ils avaient essayé de faire une partie à eux en faisant de l'hybride. Le principe de deux assistants vocaux qui se passent la main. Ils n'ont finalement pas rencontré le public. Je pense qu'ils ont essayé, ils ont tenté le *business model* en disant "nous on est Orange et fournisseurs de contenu, on a une base installée de clients Orange". Ils n'avaient pas suffisamment, je pense, de *business model* pour le faire.

Il faut savoir que tous ces acteurs-là qui se lancent dans ces assistants vocaux, ne font pas d'argent sur la vente des enceintes, ne font pas d'argent sur l'accès au service. C'est surtout beaucoup d'investissements, beaucoup de ventes à perte, de hardware quand ils lancent leurs enceintes. Mais le *business model* est derrière. C'est sur les ventes générées, la modélisation des requêtes et les données, etc. Donc ce n'est pas n'importe qui qui peut avoir ce *business model*.

Encore une fois, est-ce-que les assistants vocaux vont s'inviter dans les voitures ? Je dirais que ça va beaucoup dépendre du *business model*. À l'époque où les constructeurs m'avaient contactée, c'était aussi un problème de *business model*. Les constructeurs se disaient « moi j'ai un souci, ce que je vends, c'est one shot. Je vends une voiture et je ne vends pas un service avec un abonnement ». Aujourd'hui, de plus en plus de voitures sont vendues en leasing donc peut-être que ça va changer. Mais ce n'est pas encore si généralisé que ça donc c'est aussi des grandes questions. À mon avis, suivre ce secteur-là, c'est pas seulement suivre les technologies, c'est aussi suivre les *business models*.

Justement au sujet des business models, ça serait peut-être intéressant dans notre mémoire de voir aussi le point de vue des constructeurs automobiles.

Oui, si vous parlez avec quelqu'un de Faurecia, Soledad par exemple est spécialisée en ergonomie, je ne sais pas si elle saura vous parler du *business model* mais sinon elle pourra peut-être vous mettre en contact au sein de Faurecia. En tant que grand équipementier, normalement, ils connaissent très bien les *business model* des constructeurs.

Ok merci. Tout à l'heure vous parliez de barrières à l'entrée au niveau technologique, on voulait aussi vous demander du point de vue consommateur, où est ce qu'ils se positionnent en termes d'adoption des assistants vocaux dans les voitures ?

C'est une bonne question, je ne sais pas quelle est la réponse. Le débat a commencé à partir du moment où les assistants vocaux de type Siri se sont développés. Les constructeurs se sont demandé s'ils devaient développer leur assistant vocal intégré dans la voiture ou bien est ce que les gens allaient rester sur leur smartphone dans la voiture. Je pense que c'est un grand débat.

Ce n'est pas idiot l'idée de faire un espèce d'hybride. On a deux assistants vocaux : celui sur la voiture et celui qu'on utilise généralement. C'est un grand débat que j'ai vu arriver puisque j'ai vraiment croisé de près ce marché. Maintenant je ne le connais plus si bien aujourd'hui. Lors de mes débuts chez Nuance, dans les années 2000, les assistants vocaux n'existaient pas. Bien sûr, mettre de l'interaction vocale dans une voiture, c'était obligatoirement de la technologie embarquée, spécifique, etc. C'est ce qu'on faisait chez nuance en embarquant des puces de reconnaissance vocale dans les habitacles avec plein de travaux liés à la captation audio. On se posait des questions, « est-ce qu'il faut un ou deux micros ? ». C'est une autre raison pour laquelle au début ce n'était que les habitacles haut de gamme qui en étaient équipés, outre le côté silencieux de l'habitacle dans les modèles haut de gamme. C'était aussi le fait que le constructeur pouvait se permettre dans ce type de modèle avec deux micros. Or, aujourd'hui, il faut savoir que dans tout ce qui est reconnaissance vocale ambiante, c'est-à-dire enceintes connectées, reconnaissance en voiture, situation mains libres, il faut normalement deux micros. Si vous n'avez pas au moins deux micros, c'est quasiment impossible d'avoir une reco vocale de haute qualité. Alors pourquoi deux micros voire plus ?

D'une part, ça permet de faire une antenne, donc de localiser d'où vient la source et bien se caler sur la source de la voix pour récupérer le son qui vient uniquement de cette source-là, à l'exclusion de toutes les autres. En effet, l'habitacle peut faire toutes sortes de bruit. Même sur votre téléphone, vous avez deux micros. Là, ce n'est pas pour faire antenne parce que normalement vous parlez très près de votre *smartphone* donc il n'a pas forcément besoin d'un effet d'antenne. Le deuxième microphone se trouve derrière le capot et il permet de faire de l'annulation d'écho. Quand vous parlez à votre *smartphone*, il capte ce qui vient de votre côté et il capte ce qui vient de l'autre côté de la pièce pour l'annuler. C'est très intéressant parce que si vous êtes dans une pièce avec un peu de bruit ou même dans une pièce où vous avez le sentiment d'être au calme, il y a toujours des petits bruits. Il peut y avoir des voitures qui passent, il peut y avoir des personnes qui circulent, un son, etc. Donc c'est cette annulation d'écho est très importante et permet d'avoir des qualités de reco très supérieures. C'est juste pas jouable sans ça.

Dans les débuts de l'assistant vocal en voiture, il y avait toutes ces questions-là sur l'antenne de micro. Comment est-ce qu'on annule le bruit de l'habitacle ? Ça faisait beaucoup de questions de hardware à embarquer dans l'habitacle parce que la question ne se posait pas tellement de dire « est-ce qu'on fait ça dans le cloud ? », c'était tout embarqué. Ensuite avec les assistants vocaux, on s'est posé les questions : « qu'est-ce qu'on fait ? », « est-ce qu'on y va, est ce qu'on n'y va pas ? ». Donc on va faire un assistant spécifique, je pense que la question continue à se poser aujourd'hui. Est-ce que les utilisateurs vont plutôt aujourd'hui utiliser leur *smartphone* qu'ils ont mis sur un petit support ? C'est intéressant de poser la question avec le GPS parce qu'il y a des débats qui sont communs entre les deux. C'est à dire qu'à une époque, c'était du haut de gamme d'avoir un GPS dans une voiture. Ensuite, les gens achetaient le leur, puis ça a été intégré et embarqué dans la voiture. Aujourd'hui les voitures ont souvent le GPS mais de plus en plus les conducteurs utilisent Waze ou leur *smartphone*, plutôt que le GPS fourni avec l'électronique de la voiture. Je pense que c'est un petit peu le même débat avec l'assistant vocal.

Après c'est juste la question, est-ce que l'utilisateur en voiture veut interagir de façon spécifique avec sa voiture ou est-ce que quand il parle il veut les fonctions que lui offre l'assistant vocal (c'est à dire il veut passer un appel ou composer un numéro sans avoir à utiliser ses mains ? Aujourd'hui, la plupart des voitures ont le système de connexion Bluetooth du téléphone. Vous êtes quand même obligé d'appuyer sur certaines choses et de regarder un écran pour composer

un appel. Même si c'est plus ou moins bien fait, ça vous distrait de la route, c'est très clair. Donc, pouvoir tout faire à la voix, sans retirer du tout son regard de la route, c'est intéressant. C'est typiquement offert par les assistants vocaux depuis très très longtemps, même avant l'existence de Siri. C'était une des premières fonctions vocales que vous aviez sur les smartphones : la composition à la voix des numéros et d'appels des contacts. Donc cette fonction-là, c'est sûr qu'elle est très utile mais il est inutile d'avoir un assistant vocal spécifique à la voiture pour faire ça.

Est-ce que finalement l'assistant vocal du smartphone le fait très bien ? Est-ce que la personne va vraiment utiliser la voix ? Je sais pas vous dire. Après je peux vous donner mon opinion en tant que conductrice mais ça ne sera pas en tant qu'experte. Je vous parlerai des voitures de marques, peut-être parce que quand vous avez une voiture que vous ne connaissez pas, ça peut être intéressant. Je vous dis ça parce que ça m'est arrivé, ma voiture est au garage donc là j'ai une voiture de remplacement en attendant. Donc c'est une voiture que je ne connais pas bien, les voitures sont toujours faites plus ou moins pareil. On s'adapte du changement de vitesse à l'automatique. Ça m'a un peu saoulé au début mais bon quand on a appris à conduire avec le changement de vitesses, qu'on sait changer les vitesses, ça revient. Mais hier je me suis retrouvée sous la pluie, j'ai dû déclencher l'essuie-glace arrière et là, il y a une petite variante donc je me suis énervée. Heureusement, c'était pas si différent que ça mais j'ai mis un petit peu de temps avant de comprendre. Puis il y a une autre fonction qui m'a vraiment ennuyée parce que ça commençait tout doucement à s'embuer. Là c'est très dangereux, je ne voyais plus rien et il y a des voitures où ce n'est pas du tout intuitif. Je suis habituée à des voitures familiales où vous avez un tableau de bord qui est assez simplifié. J'avais oublié que dans beaucoup de voitures, c'est pas simplifié comme ça. Il faut manipuler au moins trois petits boutons pour envoyer l'air vers le haut. Je me suis même dit que c'était de la non-assistance à personne en danger de laisser ce choix non ergonomique.

Le temps que je comprenne, ça a certainement détourné mon attention. S'il y avait eu un assistant vocal, j'aurais juste dit : « enclenche le désembuage ». Donc c'est vrai que pour des voitures de location, pour des gens qui ne les connaissent pas, ça peut être intéressant et sauver des vies. Maintenant quand vous êtes habitué à une voiture, en fait c'est complètement un réflexe. Mais ce n'est pas du tout un avis d'expert donc à prendre avec des pincettes, c'est vraiment du commerce.

Par rapport à ce que vous venez de dire, est ce que vous pensez que l'assistant vocal peut apporter une sécurité supplémentaire à la conduite ?

Théoriquement, oui. Théoriquement tout ce qui peut permettre de faire des requêtes à la voix sans enlever ses yeux de la route, c'est intéressant. Vous en avez l'intuition, je vois ça dans une de vos questions, qu'il y a un risque aussi. C'est-à-dire que parler, ça peut détourner l'attention. Je pense que ceux qui développent les assistants vocaux pour les voitures en sont très conscients. Ils savent qu'ils doivent faire très attention à la charge cognitive parce qu'au lieu d'être une aide, ça peut très vite être une vraie charge, si les choses ne se passent pas de façon intuitive et s'il faut se mettre à réfléchir et à devoir corriger. Attraper un dialogue un peu complexe, ça peut complètement détourner l'attention.

On voit bien que quand on est au téléphone en voiture, même en *bluetooth*, le fait de parler avec un interlocuteur nous projette dans un autre lieu et réduit la vision périphérique, l'attention. On peut facilement louper une sortie ou se tromper de route quand on parle. Le fait de parler est une charge cognitive, y compris avec le passager. Notre cerveau n'est pas si multitâche que ça. Il y a des tâches qui peuvent être faites plus ou moins bien en tâche de fond. Parfois il y a des tâches cognitives qui n'ont qu'un seul centre. Le cas classique, qui n'a rien avoir avec la voiture, c'est quand on est en réunion, on reçoit un SMS et on a envie de le consulter, en écoutant d'une oreille. On pense être capable de faire les deux mais on ne se rend pas compte que dès qu'on a passé la phase de reconnaissance visuelle et qu'on commence à lire le message, là c'est fini. Physiologiquement, on ne peut pas écouter une réunion tout en lisant un message ou alors c'est un entraînement. Les pilotes d'avion militaires ont des entraînements pour pouvoir entendre en même temps plusieurs voix qui arrivent dans le cockpit. Ils arrivent à faire des micro-interruptions. C'est un effort conscient de constamment interrompre son attention pour passer d'une source à l'autre. Mais c'est épuisant et ce n'est pas du tout intuitif ou naturel. On pense qu'on est capable de faire ça mais c'est faux. On peut le faire uniquement quand on utilise le cortex visuel au tout début mais ensuite c'est très vite intégré avec les zones du langage.

Du coup je ne vous donne pas une très bonne réponse mais par exemple, si j'avais pu dire à la voiture de lancer le désembuage, ça aurait servi. D'ailleurs c'est un peu ce que j'ai fait parce que j'avais une passagère.

Au début de la composition vocale des numéros, quand on composait un numéro chiffre à chiffre, au début des années 2000, j'ai découvert un truc. J'étais dans le monde de la reco de parole et c'est pour ça que j'ai compris quel était le problème. Au début, ça se plantait tout le temps. J'avais beau dicter le numéro, le système s'arrêtait avant que j'ai fini ma dictée. C'était très frustrant, il s'arrêtait au milieu du numéro et il disait « je compose le ... » et citait les six premiers chiffres que je venais de dicter. Je suis dans le monde de la reco de parole donc j'ai vite compris qu'en fait c'était un problème de *endpointing* parce que moi je travaillais sur ces sujets-là donc je connaissais bien tous les problèmes de *end* et comment fonctionnait le système pour comprendre que vous avez fini de parler. C'est quelque chose de très important dans l'ergonomie d'un dialogue vocal, c'est cette question de comprendre à quel moment on a fini de parler. Même nous, les humains, on a du mal. Parfois, on se coupe la parole sans l'avoir voulu et donc c'est très compliqué d'évaluer quand est-ce que la personne a fini sa prise de parole et qu'il faut passer à la suite. Donc on utilisait en reco de parole, les silences. Il y a toutes sortes d'indicateurs qui peuvent aider le système à comprendre que l'énoncé est terminé et qu'il faut passer à la suite. Je connaissais la question de *endpointing*. Le système décidait d'interrompre son écoute dès qu'il avait un tout petit silence. Je pense qu'ils avaient fait ça parce que le fait de réduire le temps de réponse après un silence, ça permet d'être super rapide et d'avoir un dialogue super fluide. À peine la personne a fini de parler que la réponse arrive. Ça doit être génial en démo mais dans la vraie vie c'est une catastrophe. Ça veut dire que dès que la personne reprend son souffle, le système l'interrompt.

J'avais compris qu'il avait calé sa pause de fin de parole hyper courte, sur une demi-seconde. Sachant ça, quand je dictais un numéro, je me concentrais, je prenais mon souffle et je disais à mon numéro en faisant bien attention de ne pas laisser le moindre espace de silence entre deux nombres. C'était une tâche qui ne demandait aucune charge cognitive. Mais c'était devenu un effort, qui demandait une grande concentration. Ça pouvait, pendant ce temps-là, distraire de la conduite.

Voilà donc je ne sais pas faire l'évaluation bénéfice-risque, ça dépend beaucoup de l'ergonomie. L'ergonomie n'est pas simple. En effet, là ce qui semblait être un bon choix ergonomique, de caler la fin de parole sur une pause très courte, de façon à avoir un dialogue fluide, était finalement une très mauvaise idée. Chez Nuance, on savait depuis longtemps que dans ce genre de cas, il faut vraiment laisser les pauses maximales et détecter la fin de parole

avec un autre indicateur. Il faut détecter la fin de parole par une indication sémantique, c'est-à-dire de comprendre que ce qui vient d'être dicté est un numéro. Il y a donc de fortes chances que la personne ait fini de dicter. Il fallait utiliser le contexte et la sémantique.

On avait justement une question à laquelle vous avez déjà un peu répondu avec cet exemple, quelles sont les demandes qui restent encore trop complexes à traiter pour un assistant vocal ? On avait vu que le cumul de plusieurs paramètres pouvait être trop complexe.

Ça m'intéresse, pourquoi vous pensez que c'est quelque chose qui mettra l'assistant en difficulté ?

On se dit que pour le moment, l'assistant vocal n'est peut-être pas encore assez développé pour combiner plusieurs paramètres. Il est peut-être seulement adapté à combiner deux paramètres.

Vous avez raison, ça peut être une limitation. Il faut bien comprendre où sera la limitation. C'est intéressant, ça me fait penser à un déploiement qu'on a fait récemment dans le domaine du *call bot* ou de l'accueil vocal en téléphonie mais c'est un petit peu la même chose. Vous, ce que vous voyez, c'est le résultat final, c'est cette interaction avec un système. Le système en fait il a plusieurs couches technologiques : la brique de *speech-to-text*, la brique d'interprétation du langage. J'ai peut-être ma vision trop biaisée sur la couche technologique.

Quand vous dites qu'ils ne sont pas encore prêts, ça peut être lié à des limites à différents niveaux. Ça pourrait être la première couche de reco vocale. C'est-à-dire que si le moteur de reconnaissance vocale a des difficultés à transcrire la phrase, qui est dite en langage naturel, ça peut être une première limite. Notez que la limite peut être de deux ordres. J'ai vu deux technologies évoluer et converger mais c'est des technologies qui avaient même deux noms différents à une époque. On parlait d'une part de reco vocale, ASR (Automatic Speech Recognition) et on parlait d'autre part de STT (*speech-to-text*), transcription voix-texte. Pourquoi c'était deux technologies différentes ? Parce que l'ASR, pendant longtemps, c'était vraiment la reconnaissance vocale guidée. C'est-à-dire qu'on va faire des grammaires, qui vont reconnaître toutes les phrases possibles que va dire une personne. Ensuite, on va énumérer les constructions de phrase et décrire de façon experte les constructions possibles de

phrases en essayant d'être le plus exhaustif possible dans le contexte de l'interaction. Donc une demande de restaurant peut être décrite dans une grammaire.

Quand je suis passée chez BertinIT, c'était carrément le monde du speech-to-text. Là il s'agit de transcrire la voix et ce ne sont pas forcément des gens qui parlent pour retranscrire. Ça peut être des émissions de radiotélévision qui sont diffusées. Les gens qui parlent à la télé pour que le moteur de reconnaissance vocale comprenne ce qu'ils disent. Ou alors vous retranscrivez des conversations agents-clients dans un centre de contact. C'est pareil, les gens se parlent entre eux, ils ne parlent pas pour qu'un moteur les comprenne. Ces prises de parole sont très différentes des prises de parole que vous avez quand vous êtes en interaction vocale avec une machine. On appelle ça des dialogues spontanés humain-humain. Là pour transcrire ces conversations, ce ne sont pas les mêmes technologies. Ce sont des modèles statistiques.

Aujourd'hui il a beaucoup de deep learning. Imaginez que vous utilisez une technologie de premier ordre, plutôt dans le monde de la reconnaissance vocale, peut-être avec un peu de statistiques, ce n'est pas un modèle qui est fait pour transcrire n'importe quelle prise de parole. Imaginez que c'est un moteur assez spécialisé de reconnaissance vocale, la première difficulté que vous allez avoir c'est que tant que vous respectez soit la grammaire qui a été écrite, soit vous êtes proches des exemples sur lesquels ont été appris le modèle de statistique de langage, ça va fonctionner. Si ça a été prévu de dire « je ne veux pas un MacDo, je ne veux pas un Quick et qui est à moins de 2km », ça passera la première étape. Maintenant aujourd'hui, on a ces moteurs qui sont très puissants. À priori, ils vont pouvoir reconnaître ce genre de choses. Il y a un certain nombre de moteurs qui ne vont pas être mis en défaut par ce genre de choses.

Après, la deuxième étape, c'est la compréhension de ce qui a été dit. Là, il y a la brique de NLU⁸⁷ (Natural Language Understanding) donc il faut qu'elle soit conçue de façon à être capable d'interpréter. C'est le même problème : est-ce que c'est des règles expertes ? Est-ce que c'est statistique ou neuronal ? Est-ce qu'il y a suffisamment de liberté pour interpréter des choses qui n'ont pas été prévues ? C'est le même débat qui va se poser sur la brique suivante.

Ensuite, par rapport à la brique de reco vocale, ça va être reconnu à 95%. Mais parfois ça ne suffit pas, ce n'est pas suffisamment précis sur les demandes importantes. Peut-être que ça va

⁸⁷ NLU, voir glossaire n°32.

très bien transcrire les petits mots et la phrase globale mais si c'est sur une information critique, c'est ennuyeux. Ensuite, il y a la couche d'après. Même si la reco vocale a marché, même si le NLU a marché, si derrière le système back-end n'a pas l'information ou ne sait pas traiter ce genre de requête, ça ne fonctionnera pas. Il existe un autre problème : celui de la base de connaissance de l'application et les types de requêtes qu'on peut lui demander.

Votre intuition est bonne de dire que c'est quelque chose qui va être difficile pour l'assistant vocal mais il faut regarder à quelle étape ça bloque. Parfois, c'est assez frustrant. Par exemple, récemment, le moteur a été déployé pour une grande banque. Il sert d'accueil vocal pour les appelants et c'est un moteur qui a un très grand vocabulaire. Il est utilisé pour transcrire des conversations humain-humain dans des centres de contact. À fortiori, il peut comprendre la demande initiale d'un appelant mais le dialogue qui a été construit autour est un dialogue hyper sécurisé. Ce dialogue fait en sorte que l'appelant est guidé et doit répondre par oui ou non. Ça donne l'impression d'un système très guidé, qui utilise une technologie ancienne et qui n'est pas capable de reconnaître des expressions naturelles. Ce n'est pas le cas, c'est plutôt comme ça que c'est construit. Parfois, les différentes couches s'interconnectent et quand ça ne fonctionne pas, on ne sait pas de quelle couche ça vient. Très spontanément, vous pensez que ça vient de l'une mais peut-être que ça vient de l'autre. Moi j'avais donné le numéro à des clients qui m'ont dit que le moteur était très limité. J'ai dû expliquer que ce qu'ils entendaient c'était pas le moteur, c'était le choix de construction du dialogue. Ça n'exprime pas la capacité du moteur. Spontanément, la personne avait attribué la limitation au moteur.

Quand vous avez un système qui ne comprend pas votre requête, c'est très difficile de savoir à quel niveau ça bloque et de savoir de quelle couche ça vient. Par rapport au votre, je dirais qu'au niveau des briques technologiques, on a tout ce qui faut. Il n'y a pas de raison que ça ne se fasse pas mais ça peut être le 2% d'erreur qui pèse. Mais dans l'automobile, je ne suis pas la bonne personne pour vous dire ce que l'on peut faire de façon fiable ou pas.

Il y a encore tout à fait l'utilité d'avoir l'assistance vocale en voiture, la voiture autonome on n'y est pas encore. Les débuts que j'ai connu dans les applis vocales avec interface vocale en voiture, c'était le premier contrat avec General Motors qu'on avait signé chez Nuance. Entre ce temps-là et aujourd'hui, ça fait presque 20 ans et il y a eu énormément de progrès dans la technologie, comme il y en a rarement. Le Deep Learning dans les apprentissages profonds a

vraiment révolutionné la technologie et les capacités de transcription vocale dans la reconnaissance vocale mais aussi de compréhension du langage. Maintenant aussi ça révolutionne la synthèse vocale. C'est une période de bond technologique. Pourtant moi j'ai le sentiment que ça n'a pas évolué tant que ça. L'équipement vocal en voiture, ça n'a pas fait un bond énorme alors que ça aurait pu. Donc peut être que c'est parce qu'il n'y a pas de *business model*. C'est compliqué pour les constructeurs. Ils ne sont pas habitués à des *business models* de cloud, d'abonnement ni de valorisation de la data ou vente de contenu/ de flux.

Donc ce n'est pas compatible avec ces services qui sont à priori très liés à ces modèles-là. Les grandes sociétés qui investissent dedans sont dans ces modèles de cloud, de service d'abonnement, de data, contenu, etc. Puis, peut-être aussi qu'au fond les gens utilisent leur assistant vocal sur smartphone et donc il n'y a pas eu de gros besoin de développer l'assistant vocal spécifique automobile. Voilà, c'est juste des hypothèses. Pourquoi ça ne s'est pas développé plus que ça ? J'ai le sentiment que ça ne s'est pas tant développé que ça par rapport aux 20 années qui se sont écoulées entre les premiers contrats et maintenant.

Merci pour cette réflexion. Je pense qu'on n'avait pas beaucoup d'autres questions. Désolé, on a un peu dépassé le timing. Je ne sais pas si vous avez encore le temps ?

Oui c'est vrai que là j'ai plus beaucoup de temps. Est-ce que vous pouvez m'envoyer ce que vous allez exploiter dans votre mémoire ? S'il y a un compte rendu, une citation, si vous pouviez me les transmettre préalablement. Et dites-moi si vous voulez que je vous aide à contacter, que ce soit Ahmed Bouzid ou Soledad Bedu, ils peuvent avoir une idée plus précise que moi sur le sujet.

Ah oui, merci. Ça serait très intéressant. Un tout grand merci en tout cas, merci de nous avoir octroyé du temps.

Je vous remercie, n'hésitez pas à rester en contact et m'envoyer parce que ça m'intéresserait aussi de voir votre mémoire. Très beau succès dans votre travail et bonne continuation.

Merci beaucoup, bonne journée.

3. Interview de Jean-François Bonastre

Bonjour.

Bonjour, voilà, je suis à vous. Pas pour très longtemps, mais une bonne demi-heure sans souci après ça sera un peu plus difficile.

Ok super, tout d'abord merci d'avoir accepté de participer à cet entretien. Pour se présenter brièvement, on est toutes les 2 étudiantes en dernière année de master en sciences de gestion à Louvain-La-Neuve. Notre mémoire a pour objectif de définir la valeur ajoutée de l'assistance vocale dans 3 cas d'usage qu'on a choisis. C'est l'aide à la conduite, la traduction et les services de secours à distance.

Donc nous, on vous a contacté pour un peu mieux comprendre tout ce que représente l'intelligence artificielle, les assistants vocaux, mais surtout pour avoir votre point de vue sur le premier cas qu'on étudie, qui est l'aide à la conduite. Donc voilà, on aimerait un peu parler des assistants vocaux dont les voitures sont équipées.

Je peux vous laisser la parole pour aussi vous présenter en quelques mots, votre parcours et peut être votre expertise en matière d'assistants vocaux.

OK. Moi, je suis prof en informatique à l'Université d'Avignon. Je travaille sur le traitement automatique de la parole depuis mon master, mon DEA⁸⁸ à l'époque, plutôt dans tout ce qui concerne les informations non linguistiques : l'identité des personnes, le stress. J'ai pas mal travaillé sur les centres de secours, par exemple, l'estimation de la charge cognitive, la langue parlée, etc.

Je fais aussi de la parole au sens plutôt reconnaissance de *wake-up word* ou d'événements sonores et des choses comme celles- là. Là on rejoint la voiture, la commande vocale en petit vocabulaire, des choses très interactives.

⁸⁸ DEA (Diplôme d'études approfondies) est un ancien ancien diplôme universitaire existant en France entre 1964 et 2005, en Belgique jusqu'en 2005, et dans des pays suivant le modèle français de l'enseignement supérieur comme le Liban ou ceux du Maghreb. Il est équivalent aujourd'hui à une 2e année de master (Wikipedia, 2021).

Voilà, sur le reste, bien évidemment, je connais, j'encadre des gens sur ce qu'il y a comme par exemple des moteurs de dialogues mais ce n'est pas ma spécialité.

Je vais quand même rajouter quelque chose parce que je suis très bavard. Par conviction et un petit peu par hasard scientifique, il se trouve que j'ai été confronté très tôt au problème d'éthique dans ce genre d'application, avec une assez grosse activité dans les tribunaux pour expliquer les dangers de l'utilisation de la technologie dans certains cas, en particulier dans l'identification des personnes, par la voix. Du coup, j'ai un attachement assez fort à tout ce qui est protection, *privacy protection*. J'ai pas mal d'écoute aux réticences actuelles dans l'usage. Mais je pense qu'on en reparlera.

Ok. Je pense qu'on peut commencer dans les questions en rapport avec les voitures. Donc peut-être la première question : selon vous, quelles sont les avancées actuelles des voitures dotées d'une assistance vocale ? Pour le moment où en sommes-nous au niveau assistant vocal dans les voitures ?

Moi je trouve que ça ne marche pas très bien. Je suis personnellement un utilisateur récent, depuis quelques années, de solutions standards type Android. Je trouve que derrière l'aspect gadget, très convaincant au départ, c'est quand même assez pratique pour passer un coup de téléphone, il faut le reconnaître. Mais pour un vrai usage, il y a des gros problèmes d'ergonomie et de performance. Malgré le fait que la principale limitation soit qu'on est en *push button* et non pas en *micro ouvert*. Je n'ai pas trouvé qu'il y ait eu des évolutions depuis 4-5 ans. Ça stagne pas mal. Maintenant je n'ai pas installé d'applis super sophistiquées ou des choses spécifiques.

Je connais bien les problèmes. Je ne sais pas si vous avez listé certainement les problèmes dans les voitures ? Nous on est dans le sud, on a souvent les fenêtres ouvertes et si vous mettez de la musique, si vous ouvrez les fenêtres, si la clim est forte, etc., vous pouvez oublier l'usage à l'heure actuelle.

Vous avez dit que c'était un gadget assez convaincant, donc est-ce que vous considérez que ça peut tout de même apporter une plus-value ou ça reste un gadget ? Est-ce que ça pourrait devenir une innovation indispensable ?

Non, non, je suis plutôt dans la 2ème option. Un gadget, je suis un peu méchant parce que j'utilise quand même réellement pour 2 applications : pour appeler quelqu'un, téléphoner et également pour rentrer ma destination dans mon GPS. Avec quelques efforts et l'usage de l'écran tactile pour lever quelques ambiguïtés, c'est quand même déjà utilisable et utile. Je m'en sers couramment quand même. Je pense que ça peut devenir extrêmement utile.

Je conduis pas mal et je suis quand même quelqu'un de relativement occupé. J'aimerais bien pouvoir dicter des petits rapports. Ça serait pratique en conduisant. Alors est-ce que c'est une bonne idée au niveau de la qualité, de la conduite et des risques ? C'est un autre problème. En tout cas, c'est mieux que d'essayer de noter sur un bout de papier tout en conduisant. Je n'ai pas trouvé de bonne application pour l'instant qui m'aide à faire ça. On en discutait ce matin avec mes collègues, il y a des applications de type Bloc-notes mais qui ne permettent pas de modifier, qui ne permettent pas de réécouter facilement. Enfin, il n'y a pas d'application interactive correcte, qui permette de faire ça. Ma limitation est plutôt là. Je pense que ça serait très utile. Maintenant peut être que c'est un choix et que la régulation fera qu'on n'aura pas le droit d'utiliser ce genre d'utilisation parce que ce genre d'applications cause des dangers.

Donc on voit que vous pouvez être un peu réticent à l'adoption des assistants dans les voitures. Quelles sont, selon vous, les autres raisons principales de la réticence d'utilisation dans la voiture ?

Après il y a des réticences techniques parce que certains éléments ne marchent pas. Il faudrait mieux encadrer l'usage sur ça. Je pense qu'à l'heure actuelle, ce qui nous bloque, c'est qu'on ne sait pas vraiment pourquoi on va mettre un assistant. Je vais expliquer : par exemple, j'ai travaillé récemment avec une société, qui travaille dans ce domaine. Il y a deux grandes applications visées, en dehors de l'aide à la conduite, c'est gérer les fenêtres et l'autoradio. Quelque part, j'ai des boutons au volant quand je conduis et je ne suis pas certain que ça soit plus facile de dire "ouvre la fenêtre gauche" que d'appuyer sur le bouton. Ce que je veux dire, c'est qu'en situation importante d'urgence, j'ai pas tellement besoin d'ouvrir ou fermer la fenêtre, que ça soit à la voix ou avec mes mains. Donc ce genre d'application, j'attends d'être convaincu. Mais je ne suis pas très convaincu. Pourquoi pas ? Mais alors plutôt sur le mode gadget. Je ne suis pas passionné par ça donc c'est plutôt ça mon type de limitation.

Le 2ème, qui peut être intéressant et très recherché en ce moment, c'est toutes les applications de commerce électronique par la voix, depuis la voiture. Il peut y avoir un intérêt à commander un ticket de péage particulier. Voilà quelques cas intéressants. Maintenant j'allais travailler sur cette appli où il cherchait à faire de la vraie commande, type Amazon, à la voix en conduisant. Je peux commander mon repas pour la prochaine station. Là c'est bien d'utiliser la voix mais je ne peux pas commander ma prochaine télé, mon prochain écran ordinateur en conduisant parce que j'ai besoin de voir les images.

Je pense qu'on manque d'application ciblée à l'heure actuelle. Peut-être qu'il y aura des applications en lien avec une chaîne de restauration avec la réservation de quelque chose à l'arrivée et le classique, la réservation des chambres d'hôtel. Ce n'est pas spécifique à la voiture et je ne sais pas si la qualité en conduisant va être suffisante. Là, il y a une limitation technique et là aussi il y a un intérêt. Si je suis en retard et je me rends compte qu'il faut que je trouve un endroit pour me loger en passant par Lyon. Tout en conduisant, je vais dire « réservez, une chambre d'hôtel ». J'ai pas essayé là, dans l'état actuel. J'ai pas eu l'occasion de le faire, mais pour l'avoir fait à la voix, à l'arrêt, là aussi, je peux le faire.

Après les autres réticences sont un peu ridicules. Enfin, importantes et ridicules à la fois. Il y a des réticences sur la préservation des informations personnelles. « Est ce que je vais être plus tracé, etc. ». Du moment que j'utilise Google en même temps, de toute manière je suis tracé. Peut-être que le fait d'avoir des assistants sur lesquels on travaille, qui sont entièrement embarqués et qui ne font pas sortir la voix, ce serait mieux. Ceci dit, je ne pense pas que ce soit une réticence du public pour l'instant.

Vous ne considérez pas la protection des données comme une réticence des utilisateurs ?

J'espère que ça va devenir une très grosse réticence. Pour l'instant, c'est extrêmement ambigu, même dans mon propre usage. Oui, j'ai beau désactiver autant que possible des choses quand on a envie d'utiliser une application, on accepte plein de trucs alors qu'on ne comprend rien. Moi j'insiste sur le côté de la voix qui donne des informations très personnelles, très différentes des autres. Mais si on revient dans votre sujet, de toute manière, on utilise tous ces applications-là, on a déjà tous donné notre accord. Donc le fait de l'utiliser dans la voiture ne rajoute pas un danger supplémentaire. Le plus gros danger dans la voiture, c'est le fait que la voiture trace en

termes de GPS et sache où on est, à quel moment et en permanence. Le fait d'utiliser la voix en rajoute un peu mais pas tant que ça.

Alors oui, c'est une vraie question, mais mon avis c'est qu'elle doit être traitée à un niveau plus élevé et pas seulement au niveau des assistants vocaux.

Une autre question, plus spécifique à l'assistant vocal : est-ce que vous pensez qu'il reste des demandes trop complexes à traiter pour l'assistant vocal ?

Oui, aujourd'hui et demain, oui.

En fait, on a vu dans une étude qu'il restait encore beaucoup de travail pour l'assistant vocal, pour analyser plusieurs paramètres différents. Par exemple, demander je veux trouver un fast food, qui n'est ni McDonald, ni Quick, c'est encore trop compliqué pour le l'assistant vocal.

Il y a une grosse avancée sur ce sujet. C'est vrai que ça marche dans des contextes un peu ouverts comme ça. Ça marche beaucoup. Mais en même temps, vous êtes dans un contexte de voiture où il y a du bruit, de l'ambiguïté et un contexte où vous ne voulez pas charger le conducteur. C'est un truc important et je pense qu'on a tout intérêt à avoir des assistants bien centrés sur les applis, des choses qui marchent avec des domaines restreints et ne pas leur demander de faire tout.

Mon petit doigt me dit que dans le futur, la vraie plus-value d'une société de services, sera peut-être capable de passer sur un opérateur humain pour les autres questions. Il faudra garder le contact humain. Donc, ne pas chercher à remplacer l'humain par la machine, mais faire faire par la machine les choses de base, simples. Puis, quand on a besoin de quelque chose de complexe, parce que j'ai une allergie, parce que je veux ceci ou cela, avoir accès à un opérateur humain.

Je pense que c'est une fausse piste de chercher à créer l'opérateur artificiel, qui va systématiquement remplacer l'humain.

En étudiant ce cas, on est parties sur l'intuition que l'assistant vocal en voiture pouvait apporter une aide et une sécurité supplémentaire à l'automobiliste, mais en interviewant une autre personne, on s'est rendu compte que le fait de parler, tout simplement, pouvait être une charge cognitive. Est-ce que vous pensez que ça pourrait avoir l'effet inverse au niveau de la sécurité ?

Oui, toutes les études le montrent. C'est assez évident que le fait d'avoir n'importe quelle activité cognitive, même écouter de la musique et le fait d'attirer l'attention dans la communication, occupe une partie de la capacité cognitive. Ça amène un risque. Pour balancer ça, il faut voir les actions dangereuses que l'on fait à l'heure actuelle : prendre son téléphone, appeler, tapoter sur l'écran, même sur un écran intégré, même avec un viseur tête haute. La lecture et le fait de détourner les yeux est forcément beaucoup plus dangereux en voiture que d'utiliser que la voix. Pour toutes ces choses-là, c'est évident. Après pour l'aide à la conduite elle-même, je ne sais pas s'il y a des applications à la voix où on réglerait, par exemple, le régulateur de vitesse à la voix. Moi, je ne les utiliserai pas.

Pour le moment, je préfère essayer de regarder dans ma tête la conduite avec des choses automatisées, des gestes et ce que je sais faire. Et puis finalement, les choses en plus, les petits services en plus, on va essayer d'utiliser la voix et où on va limiter cette charge.

Mon besoin serait d'avoir des applications qui me permettent de travailler en voiture, travailler en conduisant, dicter des notes et là il y a un vrai danger. Il faut savoir où on met la balance. Il y a un vrai danger avec ça.

Selon vous, ça n'apporte pas vraiment une sécurité supplémentaire d'utiliser la voix ? Il y a toujours quand même un risque et ça n'apporte pas vraiment une grosse sécurité en plus ?

Non, ça n'apporte pas de sécurité. Ça en apporte dans le sens où, à l'heure actuelle, dicter un SMS à la voix est moins dangereux que de taper sur son écran. Maintenant, légalement, on n'a pas le droit de taper des SMS en conduisant. Bon après, il faut voir le nombre de conducteurs qui le font. Pour ceux-là, dont je faisais partie, j'ai gagné en sécurité parce que je m'énerve contre Google à la voix, mais j'ai plus mon écran posé sur mon volant en essayant de taper.

Donc je pense qu'il y a eu un gain pour des actions de base comme ça. Mais il ne faut pas espérer aller beaucoup plus loin. On ne va pas mieux contrôler sa voiture parce qu'elle est à la voix.

Même dans le cas des pilotes d'avion, sur lesquels on travaille aussi, on n'est pas arrivé à ça. Alors, dans le cas des pilotes d'avion, c'est un petit peu différent parce qu'ils ont besoin de beaucoup d'informations. Donc la voix est intéressante pour sélectionner de l'information et recevoir de l'information. Recevoir de l'information à l'audio, on n'en parle pas assez, mais même dans la voiture c'est intéressant. L'assistant pourrait nous donner, par exemple, des informations sur l'itinéraire en nous prévenant qu'il va y avoir du monde dans 100 kilomètres et qu'il faudrait peut-être mieux sortir, plutôt que de mettre un écran avec des signes de danger. Là il y a un petit peu à gagner mais c'est vrai que dans la voiture, il n'y a quand même pas des messages d'alerte aussi fréquents. On ne peut pas imaginer dans la voiture un message ou d'autres qui dit sans arrêt "attention, votre distance c'est trop". Moi, je l'arrêtera. Il y a des voitures qui le font, c'est insupportable.

Ce n'est pas le cas en avion parce que dans l'avion, il y a 80 paramètres à vérifier. Et là l'usage audio est intéressant. Donc oui, pour répondre à votre question je ne pense pas que la communication vocale amène tant que ça dans la sécurisation de la conduite.

Ok. Est-ce que vous pensez que les constructeurs automobiles ont plutôt intérêt à intégrer leur propre assistant vocal dans leur voiture, ou est-ce que ça suffirait aux utilisateurs d'utiliser l'assistant vocal de leur smartphone ?

Je trouve à l'heure actuelle que l'intégration de l'assistant vocal est plutôt mal faite. Il y a un gros conflit entre le système d'information des voitures et l'assistant vocal sur la plupart des voitures. Dans ce cadre-là, il faudrait l'un ou l'autre. D'expérience, le problème des assistants sur les voitures, c'est qu'ils n'évoluent pas. Il n'y a pas de mise à jour.

Moi, je pencherais plutôt pour des assistants externes, qui soient les mêmes dans toutes les voitures. Je loue une voiture, je retrouve ma façon de faire, etc., mais avec une meilleure intégration. Je ne suis pas convaincu de l'intérêt. Les constructeurs travaillent beaucoup sur l'amélioration des conditions vocales parce qu'il faut placer les micros comme il faut. Vous savez qu'il y a plusieurs micros dans la voiture. Dans les voitures haut de gamme, jusqu'à 16

micros pour arriver à couvrir les choses. On arrive à avoir des applications qui neutralisent beaucoup mieux le bruit, etc. Donc pour cette partie acoustique bas niveau, c'est sûr que les applications spécifiques, les développements spécifiques des constructeurs sont intéressants. Pour la partie application, haut niveau comment je vais réserver mon restaurant, je trouve qu'on a tout intérêt à utiliser les applications standards qu'on utilise partout. Moi j'utiliserais plutôt celle de mon téléphone.

Ça fait penser à un autre type d'utilisation que vous n'avez peut-être pas rencontrée. Donc moi j'avais rencontré ça dans le cadre d'un projet européen. C'était la fonction interphone interne. Comment, dans une voiture de luxe, plus grande, on peut arriver à communiquer de devant, à derrière par l'usage de l'application audio, de manière facile et en gardant la spatialité. Moi je suis en train de conduire, il ne faut pas que je trouve la tête derrière pour savoir qui dit quoi. Il faut que je puisse reconnaître qui est en train de me parler. Donc c'est comment reconnaître qui parle à l'arrière, capter bien le son et le réémettre à l'avant de manière satisfaisante. Il y a des développements dans ce sens-là, qui sont intéressants aussi pour intégrer le mieux possible l'ensemble des passagers de la voiture. C'est des choses qui paraissent simples mais il y a un vrai intérêt pour éviter de forcer la voix, pour gérer ce qu'il se passe dans la voiture.

Est-ce que vous pensez que les constructeurs de voitures devraient intégrer les assistants vocaux, par exemple Google, Amazon ou Apple dans la voiture ? Ou alors justement développer leur propre assistant vocal pour chaque marque de voiture ?

Non, moi je pense que les assistants vocaux peuvent être définis de manière transversale aux marques. Je n'ai pas dit que c'était Google. J'attends une réticence forte des utilisateurs à utiliser des applications Google, Facebook et compagnie dans les prochaines années. Donc je pense qu'il y a des acteurs spécialisés de la voiture (il y en a en Europe), qui vont se développer et proposer. On est sur plusieurs choix d'applications donc ces applications n'ont pas besoin d'être dédiées à une voiture, à une marque particulière de voiture. Pas forcément mono constructeur mais plutôt une entente de constructeur. Puis 3 ou 4 solutions.

Là c'est peut-être un niveau au-dessus dans la reconnaissance vocale mais est-ce que vous pensez que la reconnaissance des émotions dans la voix pourrait être davantage utile à

l'assistant vocal ? Par exemple, dans la voiture, si on peut déceler que la personne a des signes de fatigue ou d'énervements par exemple dans sa voix ?

Dans ce sens-là, tout à fait. On travaille pour l'instant sur les pilotes d'avion dans ce cadre-là. C'est très intéressant effectivement, si vous utilisez la voix. Après il faut accepter que le système vous surveille en permanence.

Je n'aime pas le terme émotion parce qu'il est très mal traité. En réalité, les 3/4 des travaux sur les émotions, c'est de la reconnaissance d'émotion actée, dans les films et chez les acteurs. Et là, comme l'acteur, il a eu une instruction, il a joué une émotion. Effectivement, on reconnaît l'émotion. Je parle plutôt d'état émotif ou psychopathologique dans lequel on peut avoir le stress. Je vous disais, on travaille sur la charge cognitive. Avec les pilotes d'avion, on est en train d'essayer d'estimer le niveau d'oxygénation du cerveau. On est très loin des résultats. Enfin, ce n'est pas tout à fait vrai. On a des résultats qui arrivent à corrélérer la voix aux instruments qu'on place directement sur le cerveau. Mais, on en n'est pas à se passer les instruments. Il y a des éléments qui arrivent dans ce sens-là. Il faudrait qu'on arrive à combiner plusieurs éléments.

Effectivement, la voix est porteuse de pas mal d'informations et déceler le sommeil, c'est fait dans les trains depuis assez longtemps. Ça pourrait se faire en voiture. Le problème, c'est : est-ce que vous allez accepter que la voiture vous demande de parler toutes les minutes ? C'est un changement culturel. Si on accepte le côté sécurité, je corrige ce que j'ai dit tout à l'heure, dans ce cadre-là, il pourrait y avoir un plus sécuritaire de l'utilisation de la voix. De la même manière que d'ajouter un éthylotest à une voiture, c'est connu pour être un élément de sécurité depuis 30 ans. On ne le fait pas parce que on n'est pas prêt à l'accepter. Jusqu'où vous avez envie d'aller pour accepter ça ? C'est peut-être votre boulot de le faire accepter aux utilisateurs avant notre boulot, de le développer.

Et est-ce que vous pensez qu'à l'avenir, ce sera une norme d'installer des assistants vocaux dans les voitures ?

Oui, ça l'est quasiment. Il n'y a pas de doute là-dessus. La gamme des possibilités de l'assistant est une vraie question jusqu'où on va aller. Mais je crois que tout conducteur a envie de pouvoir utiliser les choses. On peut lire nos messages WhatsApp en conduisant.

Ok, merci. je pense qu'on a couvert toutes les questions. Peut-être une dernière question pour le futur : qu'est-ce que vous imaginez comme futur pour les assistants dans les voitures ? Vers quoi on se dirige ?

Mon sentiment est qu'on va aller vers une gamme de services plus restreints que ce qu'on imagine maintenant (comme je le disais tout à l'heure). Des services qui marchent bien et qui permettent d'avoir facilement recours à un opérateur humain derrière pour les cas particuliers.

Je pense que, petit à petit, on va se cantonner. Par exemple, en termes de commerce en voiture, on va se cantonner à réserver un restaurant, un parking ou un hôtel. Donc, il s'agit de quelque chose qui est très lié à la conduite, très ramené là-dessus. Et ne pas réserver une place de cinéma en conduisant. Voilà un service moins ouvert que ce qui est proposé actuellement en termes de marketing. Par contre, beaucoup plus robuste et fonctionne bien. En conduisant, on n'a pas la même patience qu'à l'arrêt pour faire et refaire des choses. Vous le voyez, vous-même quand vous essayez d'appeler. Quand j'essaie d'appeler quelqu'un, c'est que c'est important. Au bout du 2ème échec, je prends mon téléphone et puis je vais chercher le contact. À l'avenir, il vaudrait mieux, de mon point de vue, renforcer une gamme de services bien centrée sur la conduite, que de penser qu'on va utiliser les mêmes assistants sur le téléphone. Petite réflexion mais je n'ai pas beaucoup plus.

On n'a pas encore parlé d'identification. Évidemment, la plupart des constructeurs et la plupart des fournisseurs de services cherchent à ajouter l'authentification par la voix. Justement pour payer les micro services et payer les petites choses observées, etc. C'est quelque chose qui est déjà dispo, qui arrive et qui par la voix, peut donner le même renforcement que quand on l'utilise ailleurs.

Évidemment, vous avez un mot de passe. La voix est utilisée en tant que mot de passe mais pour ajouter vos connaissances là-dessus, on peut facilement, vous demandez votre date de naissance sans faire un gros effort de conduite. Vous pourriez donner votre date de naissance

pour renforcer la sécurité. Ça, c'est un service qui arrive et qui, quand il sera bien intégré sur une gamme de possibilités restreintes où l'on sait que quand on fait une réservation, il y a peut-être un blocage de 10€ mais ça dépasse pas ça. Ça vous permet, à la voix, de valider sans prendre de gros risques. Et ça marchera bien, oui. Voilà, c'est le petit ajout de la fin pour revenir sur mon thème de prédilection.

Ok. Super, merci beaucoup. Je ne sais pas si vous avez quelque chose à ajouter ?

Non, j'aimerais juste que vous me teniez au courant. Alors d'abord n'hésitez pas, si vous avez d'autres questions. J'ai aussi d'autres collègues ici si jamais vous aviez des questions plus techniques ou sur le dialogue. Et puis je veux bien avoir votre rapport. Évidemment, ça m'intéresse. Courage.

Bravo d'avoir choisi ce sujet, je ne sais pas comment vous êtes arrivées là, mais ça ne doit pas être si standard. C'est sûrement parce que vous connaissez quelqu'un qui travaille là-dessus.

Non, ce n'est pas ça. C'est parce que ça nous intéressait donc on a voulu faire notre mémoire sur ce sujet.

C'est super, puis ça montre que c'est un sujet qui devient important. Je vous souhaite beaucoup de réussite.

Merci beaucoup, merci pour votre contribution.

Merci, au revoir.

4. Interview de Soledad Bedu

Bonjour, désolée je suis sur mon téléphone car je n'ai pas retrouvé votre invitation avec mon ordinateur.

Bonjour, pas de soucis. Tout d'abord, merci d'avoir accepté de participer à l'interview. On va peut-être se présenter brièvement avant de commencer. Donc Alice et Pauline, on est toutes les deux en dernière année de master en sciences de gestion, on étudie à l'université de Louvain-la-Neuve en Belgique. Donc notre mémoire, il a pour objectif d'étudier la valeur ajoutée des assistants vocaux dans plusieurs cas d'usage, qui sont l'aide à la conduite, l'apprentissage d'une langue étrangère et les services de secours à distance. Donc nous, on vous a contactée pour un peu mieux comprendre tout ce qu'il y a derrière l'intelligence artificielle, les assistants vocaux et aussi avoir votre point de vue sur le premier cas qu'on étudie ; l'aide à la conduite. Donc parler des assistants vocaux dont les voitures sont dotées.

Ok super. Je peux vous demander qui vous a conseillé de me contacter ?

Oui, c'était Ariane Nabeth-Halber qui nous a recommandé de vous contacter.

Ah oui d'accord !

Voilà, donc on va peut-être vous laisser la parole pour vous présenter, votre parcours et votre lien en matière d'assistants vocaux.

Ok, du coup moi c'est Soledad Bedu. Moi je suis linguiste de formation, j'ai fait un master en ingénierie linguistique et traitement de la communication à la faculté d'Orléans. Alors c'était il y a très longtemps, c'était les tout premiers masters qui existaient. Et ensuite j'ai 15 années d'expérience dans la reconnaissance de la parole. Donc j'ai travaillé vraiment 7 ans au cœur du système en faisant des corpus d'apprentissage, en améliorant des modèles de langage, etc. Et puis après, j'ai travaillé 4 ans dans l'implémentation dialogue homme-machine avec des robots humanoïdes. Je ne sais pas si vous connaissez NAO, Pepper ?

Non.

Bref, on a vraiment travaillé avec des robots humanoïdes. Donc là j'ai appris aussi beaucoup du coup sur les assistants vocaux parce qu'on n'était plus à créer un système mais vraiment à interagir avec un humain. Donc on apprend beaucoup. En plus, c'était aux prémices. Quand je m'y suis mise, c'était le début aussi, donc on a appris beaucoup de nos erreurs. Et ensuite, là ça fait 3 ans que je travaille pour l'implémentation d'un assistant vocal dans un véhicule.

Donc nous, c'est surtout la partie qu'on appelle *infotainment*, donc tout ce qui est radio, le téléphone connecté, navigation, etc. En tant que linguiste, je suis spécialisée dans l'UX conversationnel. Donc l'UX conversationnel ça va être « quelle est la personnalité de l'assistant vocal ? », écrire les dialogues. Donc écrire les dialogues c'est savoir ce qu'on peut comprendre, comment l'assistant va répondre pour guider l'utilisateur. Finalement ce n'est pas si simple que ça d'écrire des dialogues. Et puis après il y a toutes les contraintes *software*, on travaille avec des développeurs pour savoir ce qui est possible et pas possible. Il y a des choses qui ne sont pas possible de mettre en place donc il va falloir trouver des stratégies, etc. Voilà.

Ok. C'est déjà très intéressant. Du coup on peut peut-être commencer vraiment dans le vif des questions plus en rapport avec les assistants dans les voitures. Donc, notre première question, c'est : quelles sont les avancées actuelles que vous connaissez dans les voitures qui sont votées d'un assistant vocal ?

Alors les avancées actuelles. Alors au départ, il faut savoir que dans le monde de l'automobile, ce sont les premiers à avoir utilisé la reconnaissance de la parole. Alors au début, ce sont des commandes très simples. Par exemple « appeler » et puis on donne le nom de son contact. Ce n'était pas naturel du tout, mais ils l'ont fait très tôt. La raison pour laquelle ils l'ont fait très tôt, c'est qu'il y a un côté sécuritaire dans la voiture. C'est que dès qu'on veut faire quelque chose, ne serait-ce que changer de radio ou faire quelque chose, le problème c'est que on enlève une main du volant et surtout le regard, on le met plus droit sur la route, on le met sur l'écran du milieu ou les touches. Du coup, le premier enjeu en voiture, et notamment pour les constructeurs automobiles, c'est la sécurité du conducteur et de tous les gens qui sont dans la voiture.

Du coup, dès le départ, ils ont utilisé des commandes très simples, et alors maintenant on est sur des assistants plus naturels parce que déjà avec Google et Alexa, ça a beaucoup progressé. Google et Alexa s'y mettent aussi pas mal pour la voiture. Avant c'était beaucoup en embarqué, ça veut dire que le système était dans la voiture. Maintenant, il y a les deux. Il y a aussi pas mal en 4G. Alors l'avantage avec le cloud c'est qu'on a des meilleurs résultats. Par contre, dès qu'on ne capte plus, on n'a plus rien. Après c'est sûr qu'on n'est plus maintenant sur des commandes simples, on ne va plus dire « mettre RTL ». Comme il y a eu beaucoup d'avancées dans le domaine en général, c'est aussi le cas dans les voitures et puis les gens c'est ce qu'ils attendent. Maintenant il y en a toujours pour parler de façon pas naturelle, mais du coup le but c'est aussi maintenant d'autoriser pas que les mots-clés, c'est de dire « j'aimerais écouter la radio RTL s'il te plaît » par exemple. Donc c'est beaucoup dans la compréhension.

Et puis les avancées sont aussi sur les réponses. Parce que finalement, avant c'était très simple de répondre. Mais maintenant qu'on veut faire des dialogues naturels, c'est ce que j'expliquais tout à l'heure ; il faut mettre en place des stratégies etc., pour que l'utilisateur naturellement réponde de bonnes choses, parce qu'on va pouvoir tout comprendre quand même. La façon dont on va poser la question va influencer sur la façon dont va répondre l'utilisateur. Et puis maintenant, on a des vrais écrans aussi dans une voiture. Ce ne sont plus des touches, y a des vrais écrans. Par exemple, on se rend compte que ce qui est important c'est que tout ce qui est affiché sur l'écran, il faut le comprendre. Parce que finalement l'utilisateur souvent, il ne sait pas quoi dire (en voiture ou pas en voiture d'ailleurs) et du coup, souvent, on va regarder ce qu'il y a. Alors quand on voit quelque chose d'écrit, on a envie de le dire. Sachant qu'en plus en voiture, si la personne réfléchit trop à ce qu'elle doit dire, elle ne pourra plus se concentrer. Elle va bien avoir ses deux mains sur le volant, regard sur la route, mais finalement elle aura une charge cognitive trop importante qui ne sera pas bonne. Il faut qu'elle garde le maximum de sa charge cognitive pour la conduite.

Voilà j'espère que ça répond à votre question.

Ouais, très bien merci. Et donc vous avez parlé au niveau de la sécurité, est-ce que du coup selon vous l'assistance vocale apporte une vraie plus-value aux voitures qui en sont équipées ?

Je dirais oui. Après, il faut que les gens l'adoptent. Si les gens ne l'utilisent pas du coup ce n'est pas une plus-value. C'est si c'est bien fait, alors oui. Par contre, si c'est mal fait non, parce que, c'est bête comme réponse, mais par exemple quand j'implémente les dialogues qui iraient dans un véhicule etc., je réfléchis toujours pour que ce soit le plus simple possible. Évidemment c'est imaginer aussi tous les arbres de décision. Est-ce que j'ai toutes les infos ? S'il m'en manque une il faut bien que je puisse la récupérer, mais il faut que je le fasse de façon simple. Vous voyez tout ça, il faut que ce soit intuitif pour l'utilisateur. Si ce n'est pas intuitif, de toute façon c'est très simple, l'utilisateur n'y arrivera pas et l'utilisera pas. L'idéal serait qu'il puisse avoir un dialogue sans avoir à regarder sur l'écran, ce qui est toujours un peu difficile. Par exemple pour les adresses, on est obligés de les afficher. Il y a des listes à choisir, c'est un peu compliqué.

Mais pour moi, là où c'est sécuritaire, c'est que l'utilisateur peut faire des choses sans avoir à enlever les mains du volant ou quitter la route des yeux. Et puis après, on peut aller encore plus loin. Il y a même des assistants vocaux qui sont ce qu'on appelle « proactifs », c'est-à-dire que par exemple il va dire (le système de navigation le fait déjà, mais pas à l'oral) « il n'y a plus beaucoup d'essence. Est-ce que tu veux que je cherche une station au plus près sur la route ? ». Il y a tout ce côté-là d'aller au-devant des besoins de l'utilisateur. Alors il ne faut pas non plus aller au-devant de tous les besoins parce que sinon ça va être envahissant et ça va aussi distraire un peu la conduite. Là-dessus aussi, il y a des choses comme ça où ça va aider et la personne n'aura pas à chercher elle-même les stations finalement. L'assistant va lancer lui-même la recherche et par contre, après, c'est quand même à l'utilisateur de choisir à la fin.

Et vous parliez d'adoption, est-ce que vous pensez que les utilisateurs adoptent déjà beaucoup l'assistant vocal dans les voitures ou c'est toujours en développement ?

Alors déjà dans les voitures depuis le départ, il y a des commandes vocales très simples, donc déjà tout ceux qui l'utilisaient continuent. Et puis, maintenant, qu'on soit honnêtes, avec tout ce qui est Car Play, Android Auto etc., c'est vraiment facile. Moi j'utilise pas mal Android Auto. Par exemple, je reçois un message sur WhatsApp, j'ai juste à appuyer sur un bouton pour que ça me le lise, et ça me demande « est-ce que vous voulez répondre ? » et je dis « oui ». Et puis après je vais répondre ça, « ok, j'ai compris ça, est-ce que c'est ça ? », « oui », « ok bon je l'envoie ». Et il n'y a pas trop d'étapes. C'est très rapide et je n'ai pas été lire sur mon portable, taper ou quoi que ce soit, j'ai gardé mes mains sur le volant. Alors qu'avant justement, alors

pour le coup je ne vais pas faire la bonne élève, mais je ne prenais jamais mon portable en voiture parce que je trouve ça trop dangereux. Et donc là je réponds parce qu'on me le propose et que ce n'est pas compliqué. Bon après si ça ne marche pas je ne veux pas insister. Mais sinon je ne le ferais pas, et du coup je trouve ça assez pratique. Et je pense que Android Auto et Car Play sont des choses qui sont faites de façon très simple pour l'utilisateur, et du coup ça va amener les gens à utiliser, et puis après, souvent les constructeurs eux-mêmes font aussi des choses. Par exemple Mercedes, c'est eux les plus avancés, ils ont un système qui s'appelle MBUX, je ne sais pas si vous connaissez ?

Non.

Allez voir, il y a plein de vidéos, c'est vraiment le système le plus avancé. C'est vraiment le système de Mercedes. Pas dans toutes les voitures, au départ c'était que dans la classe A maintenant, ils ont étendu. Donc c'est MBUX. C'est un vrai assistant qui prend en charge plein de choses. Y a des choses moins bien, mais bon on apprend aussi. Il faut savoir que c'est un domaine dans lequel on apprend encore, il est encore en construction. Par exemple ils font une option écrire un SMS. Alors moi je l'ai testée il y a 2-3 ans donc je suppose que ça a évolué depuis, mais j'ai dit jamais j'écirai un SMS par ça, c'était trop compliqué. Mais j'imagine que depuis ça a évolué, que même eux ont appris de leurs erreurs.

Et comme c'était dans les voitures qu'on a connu les premiers assistants vocaux, au début, ça pouvait être considéré un peu comme gadget. Est-ce que vous pensez qu'on est toujours au stade de gadget où ça va devenir une innovation de plus en plus indispensable ?

Alors je pense que le mot « indispensable » est peut-être fort parce que on peut conduire sans assistant vocal. Le but premier d'une voiture est qu'elle nous emmène d'un point A à un point B. Par contre, pour moi, c'est un plus pour la conduite. Comme c'est un domaine en plein essor, hors véhicule, beaucoup d'applications en reconnaissance vocale, beaucoup de personnes en Google Home ou Alexa connectent des objets. On est chez soi, on a connecté son aspirateur autonome et on dit « Ok Google, lance l'aspirateur » etc. Les gens s'habituent déjà à ça. Les gens s'habituent aussi aux applications vocales. Alors après il y a aussi des systèmes qui ne sont pas du constructeur lui-même. Mais par exemple si le constructeur décide de mettre Alexa dans

sa voiture, on peut se connecter à son compte Alexa, à sa son enceinte qui est chez soi, et puis aussi dire « Celui-ci arrive, ouvre le garage » par exemple.

Donc je pense que ce n'est pas indispensable, par contre je pense que les gens qui l'utilisent ont apprécié vu qu'ils l'utilisent aussi dans la voiture. Et puis il y a des sondages qui paraissent régulièrement, et un des cas d'usage qui ressort le plus où les gens trouvent que c'est le plus nécessaire, c'est dans la voiture. Parce que ça apporte un confort à la conduite, une aide à la conduite, qui n'est pas négligeable. Après avec les avancées Google, Alexa etc., je pense qu'on peut vraiment faire des choses d'un niveau un peu plus compliqué. Et bon c'est vrai que c'est un peu gadget d'ouvrir le garage etc., mais quand on aime ces choses-là et qu'on a tout chez soi, ça reste quand même quelque chose de pratique finalement.

Et vous pensez justement que les utilisateurs aiment bien ce genre d'écosystème ? Comme ils ont par exemple Amazon qui peut connecter la voiture avec la domotique et les objets connectés dans la maison ?

Oui, alors je pense que c'est une niche, dans le sens où tout le monde n'a pas un Amazon Alexa et tout le monde n'a pas des choses connectées. Mais quand on les a, je pense que c'est un plus de pouvoir les avoir dans les voitures. Il y a aussi des débats avec l'utilisation des données, donc beaucoup de personnes ont des craintes aussi avec Google et Amazon. Après il ne faut pas forcément passer que par eux, on peut aussi avoir un système natif. Ce qu'on appelle le système natif, c'est par exemple Mercedes avec son MBUX, donc des choses vraiment de la voiture. Là on n'aura pas forcément des choses connectées avec la maison.

Dans quelques années, si on passe par des voitures autonomes, semi-autonomes, ça peut être aussi pratique d'avoir un assistant vocal pas que pour la conduite. Il y a aussi plusieurs choses à prendre en compte. On est beaucoup pour un assistant vocal pour le conducteur, parce que la base c'est d'aider la conduite, mais on peut imaginer aussi qu'il y en ait un pour les passagers derrière. Quand il y a des tablettes, on pourrait dire « je veux regarder tel film ». Enfin voilà, on peut aussi aller beaucoup plus loin, suivant comment la voiture est équipée, et favoriser la personnalisation.

Et au niveau de la gestion des données personnelles, est-ce que c'est une grosse crainte du point de vue de l'utilisateur ?

Oui, on a bien vu quand les gens ont découvert qu'en fait Amazon et Google enregistraient finalement les données, en fait les gens n'en avaient pas conscience. Alors moi je suis mal placée parce que, comme je sais comment ça fonctionne, je le savais. Mais je comprends que quand on n'est pas dans le domaine, ce ne soit pas évident finalement. Pour moi, il faut une transparence. Moi comme je suis du domaine, je comprends l'intérêt d'avoir les données pour améliorer le système, tout dépend de ce qu'on fait de ces données-là. C'est là où il faut la transparence. Si on prend les données, on s'en sert juste pour améliorer le système et puis après on en n'a plus besoin, on les jette. Pour moi ce n'est pas un problème, mais après je suis du domaine donc je suis pas du tout objective.

Forcément, les gens ont peur parce que faut aussi avoir conscience des limites. Souvent, ils écoutent tout le temps et c'est aussi pour détecter ce qu'on appelle le *wake word*, le Hey Google, Alexa etc. Si ce n'est pas ça et qu'ils ne font rien des enregistrements, ce n'est pas grave, on s'en fiche. Par contre, s'ils écoutent les conversations, là c'est gênant. Pareil si je suis au téléphone, on peut faire des appels, par exemple si vous avez un Alexa et moi aussi, on pourrait s'appeler par Alexa. Il faut être sûr à ce moment-là qu'ils n'écoutent pas nos conversations privées. Il y a une méfiance des gens, ce qui est normal, parce que les conversations privées restent privées. Et après il faut une transparence de savoir ce qu'on fait de ces données-là.

Et d'ailleurs ici même parce qu'ils ont bien compris que les gens n'avaient pas trop apprécié, ils se sont sentis trompés en fait. Et donc c'est quelque chose qui reste toujours dans la tête des gens. Ils se demandent « Est-ce qu'il m'écoute ? ». Et c'est normal.

Est-ce qu'il y a d'autres craintes des utilisateurs, que vous connaissez, à l'égard de cette technologie ?

Il y a une crainte, mais c'est plus en utilisation. Un utilisateur qui est tout seul et utilisateur où il y a du public autour ou des gens autour, ne va pas réagir de la même façon. Les enfants eux le font très naturellement. Ils s'en fichent, ils se trompent, ce n'est pas grave, ils essayent. C'est super pour nous parce qu'ils y vont, ils ne se posent pas de questions. Tous les adultes ont une

façon d'appréhender, parce que pour eux c'est pas du tout naturel. Ce qui est normal, parler à une machine, c'est pas du tout naturel. Il y a la crainte de se tromper et d'échouer. Et elle est encore plus flagrante quand il y a du monde autour. C'est pour ça que je disais, il faut que quand on fait des dialogues ou même pour la compréhension, il faut aller assez loin. Dans les dialogues, il faut vraiment guider l'utilisateur. Déjà, on ne veut jamais que l'utilisateur échoue parce que sinon, c'est qu'il y a quelque chose qui n'est pas bon dans l'assistant, mais en plus il y a aussi cette crainte-là.

Donc on échoue, et puis parfois en plus on ne comprend pas pourquoi. On se dit « j'ai bien dit ». Les gens ne savent pas non plus comment parler donc ils vont sur-articuler, ce qui est pire parce que pour la reconnaissance de la parole ça ne marche pas. Ils vont parler plus fort, ils vont crier aussi mais il ne faut pas faire ça, mais ils ne savent pas. Il y a toute cette prise en main de l'outil qui est très compliquée parce qu'il n'y a pas vraiment de manuel, finalement. Et cette crainte de l'échec, je l'ai vue surtout quand je travaillais dans la robotique humanoïde. On a passé énormément de temps à observer des personnes à interagir avec un robot, ce qui est un peu les mêmes finalement. Enfin moi j'étais côté dialogue, donc je regarde tout ça. Et c'est un peu la même chose pour tout ce qui est Alexa, Google, ou assistants vocaux, c'est vraiment cette crainte de pas réussir, de mal dire. Souvent l'utilisateur se dit « c'est ma faute », alors que ce n'est pas la faute de l'utilisateur. Donc niveau des craintes, je pensais que c'est surtout ça.

Ok. Et autres que les craintes, quels freins sont les plus importants pour les utilisateurs à utiliser l'assistant vocal dans la voiture ?

Un des problèmes qui arrivent régulièrement, ce sont les accents. Parce qu'en reconnaissance de la parole, on gère des accents, mais pas tous. On va gérer les plus connus. Par exemple, l'anglais indien est quelque chose qui est bien reconnu. Il faut que le système l'ait aussi. Mais par exemple, moi si je dois utiliser mon anglais, même si je sais parler anglais, j'ai un accent. Même si je fais des efforts, je n'ai pas l'accent français français mais je l'ai quand même. Du coup, il y a des mots qui ne vont pas passer et il ne va pas me comprendre. Pourtant moi j'ai dit les bons mots mais on ne peut rien y faire pour l'accent. Et on ne pourra jamais couvrir tous les accents parce que, tous les accents étrangers pour une même langue, c'est compliqué pour faire des modèles à chaque fois. Il faut assez de données pour le faire donc ça, c'est un gros frein. Parce que du coup on ne veut pas non plus exclure ces personnes-là.

Mais la plupart des systèmes font plein de langues, donc on peut imaginer quand même que la plupart des langues natives, enfin la plupart non parce qu'il y en a tellement, mais par exemple, il y a des langues qui n'existent pas. Comme il y a très peu d'utilisateurs finalement il n'y a pas de système. Développer un système pour une langue, c'est beaucoup d'argent, beaucoup d'efforts, beaucoup de temps, donc s'il y a 3 utilisateurs, ce n'est pas rentable. Mais c'est quand même un des problèmes.

Et après, je pense aussi à des autres problèmes, parce que quand on parle de reconnaissance de la parole, on imagine intelligence artificielle. Et du coup, il y a des attentes très fortes des utilisateurs, ou même nous parfois, de nos clients. Parce que c'est l'intelligence artificielle, donc on imagine le truc qui réfléchit tout seul etc, mais on en n'est pas là. Donc cette attente souvent, c'est une déception d'un point de vue des utilisateurs ou parce qu'en fait on en n'est pas à l'intelligence artificielle dans le sens intelligence, ça réfléchit tout seul et ça comprend tout.

Est-ce que vous pensez à d'autres demandes qui restent encore trop complexes à traiter pour les assistants vocaux ?

Vous pensez en général ou dans la voiture ?

Les deux mais peut-être dans le contexte de la voiture oui.

Deux problématiques dans la voiture par exemple c'est tout ce qui est adresse. C'est compliqué parce qu'il y en a tellement que parfois la reconnaissance n'est pas super. On peut l'avoir du premier coup, mais on est obligé de mettre des listes parce qu'il y a beaucoup de choix, il y en a beaucoup qui se ressemblent. Souvent si je suis en Europe et que je suis en langue française, je n'ai pas pris les adresses françaises parce que je ne peux pas non plus ajouter celle des autres pays, ce serait trop de données à traiter. Ça, c'est aussi l'embarqué. Ça peut être en cloud mais ça ne marche plus quand il y a trop de données. Pareil quand on a de la musique et sur des clés USB s'il y en a trop, on ne sera pas bon parce qu'on n'arrivera pas à bien reconnaître parce qu'il y a trop de données. En fait, le système est beaucoup moins bon. Ça c'est une des problématiques.

Et après, c'est ce que je vous disais. Je pense que pour moi une des problématiques, c'est l'attente des gens par rapport au terme de l'intelligence artificielle et ce qu'on peut vraiment faire. Pour moi, c'est ça aussi. Et puis, je vois aussi des applications mal faites par exemple, ça m'irrite. Ce n'est pas grave si on ne connaît que 3 mots. Si on propose ces 3 mots-là à l'utilisateur et qu'il est reconnu à chaque fois, pour lui ça marche. Mais quand on laisse libre l'utilisateur, il ne sait pas trop où aller ou ne sait pas quoi faire. Je l'ai vécu, j'ai appelé je ne sais plus c'était la CAF ou la sécurité sociale, il me dit « que voulez-vous ? ». Même moi qui suis dans le domaine, je me suis dit « qu'est-ce que je vais raconter ? ». En plus c'est un cas particulier, donc j'ai dit plein de trucs, mais ça n'a pas marché. Du coup il me l'a fait faire 3 fois pour au final me dire « non on ne comprend pas, tapez 1 pour ça, 2 pour ça ».

Et je me suis dit que des applications comme ça, ça va encore plus freiner les utilisateurs à utiliser, parce qu'on est vraiment en situation d'échec. Et je me suis dit, moi qui suis dans le domaine, qui sais parler quand même à des assistants vocaux, qui ai une idée de ce qu'on peut dire, comment le dire, etc. Même moi je n'ai pas réussi. Donc c'est emmener les utilisateurs à l'échec. Et ça, pour moi, c'est quelque chose de pas bon du tout parce que ça renforce leur crainte.

Donc dans les freins, c'est bête mais avec une application mal faite à la voix, c'est l'échec direct et souvent ça rebute. Parce qu'on se dit « non je n'ai pas réussi pourtant je sais parler ». C'est ça qui est difficile, c'est que normalement on sait parler, c'est quelque chose d'inné, il ne faut pas des compétences particulières. Et même avec ça, on arrive à faire échouer des gens. C'est quand même dur.

Et des problèmes techniques, par exemple au niveau du endpointing, est-ce que ce genre de problème s'est résolu ? Donc, quand l'assistant vocal déclenche sa réponse avant que l'utilisateur ait fini sa phrase par exemple. Est-ce que ce genre de problème persiste ?

Ah oui, ça c'est pénible aussi. Souvent, c'est parce que l'utilisateur a fait une pause un peu trop longue et du coup on commence à *processer*. Ça passe aussi des stratégies qu'on met en place dans les assistants vocaux. Moi, je suis adepte de l'apprentissage par l'erreur. C'est-à-dire que l'on ne sait pas utiliser un assistant vocal naturellement, l'assistant vocal doit aussi apprendre et il doit apprendre l'utilisateur petit à petit. J'ai un exemple qui me vient en tête, c'est

quand je mets mon téléphone en Bluetooth, je veux appeler ma maman mais elle a 2 téléphones : le téléphone portable, téléphone maison. Du coup, si je dis « je veux appeler ma maman », il va me dire « Ok mais quel numéro ? » parce que l'on ne laisse pas l'assistant vocal choisir. Là, je vais dire « la maison ». L'assistant vocal par exemple, va toujours répondre de façon que l'utilisateur comprenne ce qu'il peut dire en une seule fois. Donc, « Ok, je vais appeler maman sur son téléphone maison ». L'idée derrière ça en fait, c'est que par répétition, en observation, il a été vu que souvent les utilisateurs miment, ce qu'ils disent à l'assistant. Ils ne savent pas quoi dire donc ils miment. Ça marche bien parce que les gens comprennent qu'ils peuvent le dire en une fois. Même s'ils ne comprennent pas la première fois, ce n'est pas grave. À force, ils vont comprendre et ils vont faire. C'est quelque chose qui aide énormément et c'est dans la conception quand on choisit les dialogues. Chaque mot est important. Il y a des mots auxquels on ne pense pas, mais ils peuvent avoir double sens, donc ça va être compliqué. Suivant le contexte, ils peuvent être un petit peu ambigu, donc il faut que ce soit très clair.

Vous pensez qu'on va vers un système où on va pouvoir parler de façon naturelle, comme si on parlait avec un humain, en parlant avec un assistant vocal ?

J'imagine que c'est possible. Après il faut voir jusqu'à quel point, parce que j'imagine qu'il y a des limites niveau calcul. Après ce n'est pas forcément la reconnaissance de la parole en elle-même qui est compliquée. Trouver les bons mots qui sont dits, ça marche bien. Il y a déjà des systèmes qui sont très bons là-dessus. Après, il y a le sens derrière, la sémantique. Par exemple, reconnaître de l'ironie, ça va être compliqué. On peut avoir tous les bons mots. Par contre, comprendre que c'est ironique, on va plus loin. Si on commence à faire des tirages très longues où on met plusieurs idées, il faut pouvoir extraire ces plusieurs idées. On avance dans ce domaine-là, mais je pense qu'il faudra encore du temps. Avoir des vraies conversations très complètes, je ne sais pas. Ce qui se fait aussi depuis quelques années, c'est qu'on ajoute une couche, ce qu'on appelle le Natural Language Understanding. Pour comprendre l'intention, il y a aussi pas mal de systèmes qui commencent à mettre la détection des émotions avec. Il faut coupler plusieurs systèmes pour arriver à quelque chose. Après pour une vraie conversation, tout ce qui a été en cours, ça a plutôt échoué. Je ne sais pas jusqu'à quel point on peut aller pour y arriver.

Il y a quelques années, Microsoft a fait un Chabot qui, à la fin, devenait raciste. Mais en fait, les données sur lesquelles ils avaient travaillé, c'était sur des données humaines. Donc le problème est toujours là, il n'y a pas une pensée négative. Il n'y a pas forcément un cerveau pour tout changer. Il y a des cas qui seront quand même toujours très difficiles.

Quand on utilise l'anaphore : il, elle, au bout d'un moment, on ne sait plus forcément à qui il fait référence. Même nous, parfois, entre humains, à l'oral, on est perdu. Donc il ne faut pas demander à un assistant vocal de comprendre ce que nous on n'arrive pas à comprendre.

Vous pensez que, à terme, quand l'assistant vocal sera capable reconnaître les émotions, ça représentera un avantage ?

Il y a des systèmes qui le font déjà. Il y a plusieurs détections d'émotion possibles. Quelquefois, il y a l'usage des mots. On se doute que si je dis un gros mot, à priori, je ne suis pas très contente. Quoique, tout dépend comment on le dit. Ça dépend de l'association des mots, par exemple « Oh putain, c'est génial ». Par contre, « putain, tu me fais chier », ce n'est pas bon signe.

Il y a aussi la voix, l'intonation etc. Et y a aussi les émotions sur le visage. Si on a la vidéo en même temps, on reconnaît aussi pas mal de choses. Si on arrive à combiner, on peut quand même aller plus loin dans les dialogues.

Dans l'aide à la conduite, si on arrivait, par exemple, à détecter le stress ou l'énervement, ça pourrait être une sécurité supplémentaire ?

Oui, ce qui se fait déjà, c'est la fatigue. Ce n'est pas une émotion, par contre, mais c'est un état. Ça se fait déjà dans certaines voitures. Ce n'est pas forcément couplé un assistant vocal. On sait que c'est l'une des premières causes d'accidents. On est fatigué, on s'endort et on se crash donc ça c'est important. Après peu importe ce que fait l'assistant, il met de la musique pour réveiller, il dit « je te conseille de t'arrêter ». Déjà le détecter, c'est important. Il y en a aussi qui font la détection du regard.

Pour les émotions, ça peut servir. On sent que quelqu'un est énervé, on peut lui proposer de faire une pause. Ça peut aussi être dans l'utilisation elle-même de l'assistant vocal. On sent

que ça ne fonctionne pas, que la personne s'énerve, l'assistante devrait peut-être s'arrêter à ce moment-là.

Vous pensez que l'assistant vocal est déjà assez proactif pour proposer des alternatives comme ça ?

Ça va dépendre de chaque système. Je ne suis pas sûre qu'il y en ait qui fasse ça, mais je pense que ce sont des choses qui sont en cours de développement, ou du moins on recherche. La reconnaissance de la parole a beaucoup progressé ces dernières années. Maintenant, on couple, on met d'autres reconnaissance et d'autres systèmes d'intelligence artificielle. Le but, c'est toujours d'aller plus loin.

Le but c'est d'aider le conducteur à avoir une conduite plus sécuritaire, plus confortable. Il y a plein de cartes que l'on peut imaginer. On peut vraiment aller très loin.

Vous disiez que Mercedes a intégré son propre assistant vocal, mais que d'autres constructeurs utilisaient Amazon ou Google. Est-ce que vous pensez que les constructeurs automobiles ont plus d'intérêt à avoir une technologie embarquée ou pas ?

Alors là, je n'ai pas de réponse directe parce que je n'ai pas trop d'avis sur la question. Je pense que ça dépend des moyens financiers de chaque constructeur. Il y a des constructeurs qui peuvent investir. Dans ce cas-là c'est super parce que c'est vraiment intégré au système. Ça fait vraiment partie de la voiture.

Peut-être que pour des plus petits constructeurs qui ont moins de moyens, ça va être compliqué d'investir autant. Donc là ça peut être intéressant d'avoir Alexa, etc. Je pense que ça va dépendre aussi de leurs clients. J'imagine que certaines marques ont des clientèles plutôt âgées qui ne sont pas très aptes à utiliser ce genre d'assistants vocaux. Donc est-ce que ça vaut le coup d'investir si les gens n'utilisent pas ?

Il y a peut-être aussi des marques plus familiales, donc il y a beaucoup de familles qui ont Google Home donc c'est peut-être mieux d'utiliser quelque chose que les gens vont avoir. Voilà,

ça dépend aussi de la clientèle. Parfois, des gros constructeurs vont avoir des clientèles différentes suivant leur gamme de voiture. Ça dépend aussi de l'image que la marque veut avoir.

Je pense qu'il y a des p des pour et des contres dans chaque. Ça dépend vraiment du constructeur de ses moyens, de sa clientèle, de l'image aussi de sa marque, etc.

Ça pourrait aussi contribuer à augmenter le monopole de Google et des GAFA en général ?

Bien sûr, c'est sûr qu'avant ils n'étaient pas sur le marché mais depuis qu'ils y sont, ça change pas mal de choses. Eux, ils ont les moyens humains et financiers de faire des outils qui fonctionnent très bien. En véhicule, quand c'est que sur le cloud, il y a quand même plein d'endroits en France où il n'y a pas la 4G. S'il n'y a pas un système de backup en embarqué, ça ne va pas marcher. MBUX par exemple, c'est sur le cloud mais ils ont un système où s'il n'y a plus de 4G, on peut quand même faire un minimum de choses en embarqué. Ça ne coupe pas complètement la reconnaissance.

L'idéal ça serait d'avoir une complémentarité entre la technologie intégrée et la technologie dans le cloud ?

Oui. Après, il faut quand même que l'utilisateur comprenne parce que les résultats ne sont pas forcément les mêmes et certaines choses ne sont pas implémentées en embarqué. Alors que d'autres le sont sur le cloud. C'est quand même compliqué de savoir pourquoi je n'ai pas le même résultat quand je dis la même chose. Et pourquoi je dis quelque chose qui n'est jamais reconnu. En fait c'est parce qu'il n'y a pas la 4G. Moi, quand je teste quelque chose, je cherche toujours les cas un peu compliqués. Je ne sais pas à quel point l'utilisateur lambda peut s'en rendre compte et comprendre que ce n'est pas lui qui l'a mal dit. C'est juste qu'on ne peut pas le comprendre s'il n'y a pas le cloud donc c'est très bien d'avoir les deux.

Moi je pense que le cloud permet une puissance de reconnaissance. La reconnaissance est quand même meilleure. Il faut vraiment rentrer dans des stratégies au niveau utilisateur pour qu'ils comprennent ce qu'il se passe s'il y a le cloud ou pas.

Une dernière question : pour vous, quel est le futur des assistants vocaux dans les voitures ?

Moi je laisse libre cours à mon imagination. Je sais que Google et Amazon ont pas mal de potentiel dedans. Après je pense qu'en voyant MBUX, ça donne envie aussi à d'autres constructeurs de le faire. Je sais qu'il y en a qui sont justement en train de faire leur propre assistant vocal. Je pense qu'au bout d'un moment il y aura forcément l'assistant vocal dans les voitures. Parfois ça reste encore en option. Dans ma voiture, je ne l'ai que par Android auto, par exemple. Si je mets mon téléphone et que je mets Android auto, je l'ai. Sinon je n'en ai pas dans la voiture. Je pense que c'est quelque chose qui va devenir un standard même si ce n'est que quelques commandes, les gens s'y habituent. Au minimum, on aura Android auto et Carplay.

Vous pensez vraiment que ça sera la norme que les constructeurs automobiles équipent leurs voitures d'un assistant vocal ?

Je pense que déjà tout le monde veut Android auto et Carplay. Il faut dire que c'est super bien fait et super pratique, donc c'est un peu normal. Je pense que, par contre, des gros constructeurs commencent déjà à investir dans leur assistant vocal. Une fois de plus, je pense qu'il y a des constructeurs qui ont les moyens de le faire. Les petits constructeurs, qui ont une petite production et une clientèle très réduite, je ne pense pas qu'ils aient les moyens.

Au niveau des demandes que les utilisateurs préfèrent, ça serait vraiment spécifique à la voiture ou alors est-ce que par exemple les demandes qu'ils peuvent faire à leur smartphone leur suffirait ?

Alors moi j'ai Android auto, donc c'est bien parce qu'il y a Spotify donc j'ai la musique. Après, il faut quand même limiter tant qu'on n'est pas en conduite autonome. La personne conduit donc faut aussi limiter les accès. Par exemple, dans Alexa pour l'automobile, ils ont restreint beaucoup de choses de Alexa parce qu'on ne va pas commencer à faire des jeux dans la voiture alors qu'on conduit. Mais ils ont étendu, par exemple sur la navigation, tout ce qui est connexion du téléphone, etc. On est quand même dans un cas particulier où la personne conduit. On veut qu'elle reste attentive à sa conduite. On pense toujours à la sécurité. Donc il faut que ce soit dédié à la voiture. J'imagine qu'avec des voitures semi-autonomes on pourra aller un peu plus loin. Il y en a qui commence à aller plus loin en intégrant des choses pour les passagers.

Il faut voir à quel point on fait un assistant personnel par siège. Un passager à l'arrière n'a pas du tout les mêmes demandes que le conducteur. Pour moi, tant qu'on n'est pas en conduite complètement autonome, il faut quand même que ce soit à la voiture. De toute façon, il faut des commandes dont la personne va servir. On peut proposer plein de choses, mais finalement s'il n'y a pas d'utilité derrière, ça n'a pas vraiment de sens. Pour le moment, c'est bête, en voulant changer la radio, j'ai déjà fait un écart sur la route par exemple. Je ne fais pas la maligne. Je me dis c'est vrai que si on peut le faire à la voix, c'est quand même super pratique. Je pourrais dire « je veux écouter, France Inter ». Finalement ça m'évitera de regarder.

On peut aller plus loin. Je pense radionavigation, mais ça peut aussi être « donne-moi l'état de mes pneus » pour vérifier la pression des pneus. On peut aller beaucoup plus loin dans les questions, mais toujours par rapport au domaine.

Par exemple, demander de déclencher les essuie-glaces, est-ce que ça ça pourrait être intéressant aussi ou pas ?

Alors, je dirais que non, parce qu'en général maintenant, ils ont des déclenchements automatiques, qui marchent plutôt bien. J'ai découvert ça en étant dans le domaine parce que ce n'est pas forcément un domaine à la base qui m'intéresse. Pour moi, les voitures sont faites pour aller d'un point A à un point B. On se rend compte qu'il y a beaucoup de technologies dans une voiture. Il y a déjà beaucoup de choses qui sont gérées par la voiture elle-même et je pense qu'il n'y a pas besoin de l'ajouter en commande vocale. Par exemple, moi j'ai l'option dépassement de ligne. Si je ne mets pas le clignotant et ça dépasse, ça me remet directement la voiture. La première fois, ça m'a fait bizarre. C'est super bien mais du coup, par exemple, la technologie dans la voiture fait déjà son travail. Ça ne sert à rien d'ajouter quelque chose. Il faut toujours réfléchir, est-ce que ça apporte vraiment quelque chose ou pas ? Il faut aussi se dire que quand on fait quelque chose à la voix, il faut que ce soit plus rapide que quand on fait *au touch* avec un écran. Je pense quand même que mettre les essuie-glaces en cliquant sur le bouton, c'est plus rapide que de demander de le faire. Surtout que c'est bien fait parce que comme c'est autour du volant, on n'a même pas besoin de regarder. Du coup, il n'y a pas l'intérêt de faire la commande vocale.

Donc l'assistant vocal sera utile quand il simplifiera des tâches et qu'il apporterait une aide qu'il fonctionnera bien ?

Voilà, il faut qu'il marche bien et que ce soit réfléchi. Parfois, je parle avec le client et je leur dis que le cas d'usage, je ne vois pas l'intérêt. Parfois ça va être trop compliqué à la voix. Ça va être trop compliqué. On va perdre l'utilisateur entre temps, il ne va plus l'utiliser et puis ça va servir à rien. Donc il faut vraiment réfléchir aussi à ce qu'il peut apporter. C'est souvent une règle qu'on se met pour tous les assistants vocaux. Il faut que ça soit plus simple qu'au *touch*. Je vous parlais du SMS dans MBUX, nous on l'a fait à l'arrêt (d'ailleurs on n'avait pas intérêt à le faire quand on conduit) et on a fini par avoir un fou rire avec ma collègue, parce qu'on se dit que ce n'était pas possible. C'est normal parce que c'est une tâche compliquée. Il faut aussi que les gens se résignent. On conduit, on n'envoie pas un SMS. Dans Android Auto avec WhatsApp c'est vachement bien fait. Mais je ne me lance pas dans des réponses compliquées. Je pense qu'il faut rester simple sinon ça ne marche pas.

Ok, merci beaucoup. Je pense qu'on a fait le tour de toutes les questions. Je ne sais pas si vous trop avez quelque chose à ajouter ?

Non, je pense que j'ai déjà beaucoup parlé.

Oui, c'était très intéressant, en tout cas merci beaucoup.

C'est avec plaisir. Je sais que quand on est étudiant, ce n'est pas facile de trouver des personnes en interview. J'ai mis du temps à répondre mais j'ai trouvé du temps parce que ce n'est jamais évident.

C'est gentil, merci beaucoup.

Merci au revoir.

Cas d'usage n°2 : Apprentissage d'une langue étrangère

1. Interview François-Xavier Fievez

Bonjour.

Bonjour. Tout d'abord, merci d'avoir accepté de participer à cet entretien, donc on va peut-être se présenter brièvement avant de poser quelques questions. Pauline et Alice, on est toutes les deux étudiantes en master en sciences de gestion à LSM à Louvain-la-Neuve.

Notre mémoire a pour objectif de définir la valeur ajoutée de l'assistance vocale dans 3 cas d'usage qu'on a choisis, donc c'est l'aide à la conduite, l'apprentissage d'une langue étrangère et les services de secours à distance. On vous a contacté pour avoir votre point de vue sur le deuxième cas qu'on étudie et donc voir si l'assistance vocale pourrait être utilisée comme un outil pédagogique dans l'apprentissage d'une langue étrangère.

Je peux vous laisser aussi vous présenter et expliquer votre parcours si vous voulez.

Oui. Alors moi j'ai fait des études en langues et littératures germaniques, donc à l'UCL. Je suis agrégé, donc j'ai un diplôme pédagogique et je travaille à l'université de Namur depuis ma sortie de l'université donc ça fait depuis 2001.

Et en fait, je travaille au sein du département de l'école des langues vivantes, donc c'est le département qui organise les cours de langues pour ce qu'on appelle les non spécialistes, donc c'est-à-dire des étudiants qui ont un ou plusieurs cours de langues dans leur cursus. Alors moi je donne principalement des cours d'anglais et depuis maintenant 5 ans j'ai aussi une tâche de coordinateur pédagogique au sein de l'école des langues vivantes. C'est une sorte de veille pédagogique et de suivi de la qualité des enseignements que l'on donne aux étudiants, notamment en faisant de l'innovation pédagogique et en faisant l'état des lieux de ce qui existe dans les nouvelles pédagogies et qui propose, en collaboration avec les facultés qui nous demandent d'organiser des cours, de monter des projets d'apprentissage des langues en fait fondamentalement.

Ok, merci. Maintenant on peut passer aux questions spécifiques à notre cas d’usage avec les assistants vocaux. Est-ce que vous connaissez les avancées actuelles mises en place en termes de assistants vocaux dans l’apprentissage d’une langue ?

Je ne peux pas dire que je connais les avancées actuelles parce que ça serait vraiment une très grande phrase. Mais depuis maintenant 3 ou 4 ans, je collabore avec une start-up belge qui a démarré d'ailleurs à l’UCL, qui s'appelle Flowchase, qui est une application de développement de la prononciation via une application mobile. Du coup, je me suis tenu au courant des différentes avancées qu'ils utilisaient, mais de là à dire que je les connais toutes, ce serait bien loin. Moi je suis pas un technicien, je suis pas un informaticien donc toutes ces avancées sont intéressantes pour moi d’un point de vue pédagogique, didactique et dans les salles de classe. Mais là, j'aurais du mal à répondre oui à cette question.

Et est-ce que vous pensez que l'assistance vocale peut apporter une aide dans l'apprentissage d'une langue étrangère ? Est-ce que ça peut rendre l'apprentissage plus qualitatif ou plus rapide peut-être ?

Pour moi, l'assistance vocale est intervenue dans cet esprit qui défend l'idée qu'une langue, ça ne s'apprend pas que pendant l'heure de cours. Donc il faut que l'étudiant ou l'étudiante puisse pratiquer, étudier, s'entraîner, s'exercer en dehors des salles de cours. Donc ce qui veut dire en totale autonomie. Et la prononciation et l'expression orale, c'est malheureusement une des compétences dans l'apprentissage d'une langue qui est très difficile à faire seul, puisqu'il y a besoin d'une correction, d'une validation, d'un feedback, etc. Donc moi je pense que dans le cadre d'un travail en autonomie sur, par exemple, un étudiant ou une étudiante qui se dit, “voilà ma prononciation, elle est quand même pas top”. Je participe au cours mais bon, quand on a 30 personnes dans un cours, si vous faites l'expression orale, tout le monde n'a pas l'occasion non plus de beaucoup s'exercer. Je trouve que le fait d'avoir un système que l'on peut utiliser chez soi, autant de fois qu'on veut, avec une qualité de correction de la prononciation, dans ce cas-ci, oui, ça c'est sûr et certain que ça me semble répondre à un besoin vraiment important dans l'apprentissage d'une langue.

Alors plus qualitatif, oui, parce que ça permet de faire, de refaire, de re-refaire et de s'entraîner autant de fois qu'on veut. J'aurais tendance à dire aussi plus rapide puisque du coup on avance plus vite. Je dirais juste à ce niveau-là, alors oui.

Donc pour vous, l'assistant vocal peut être plus utile dans le cadre d'un apprentissage autodidacte et moins dans une classe ?

Oui, en tout cas moi dans le périmètre dans lequel je l'utilise, c'est vraiment dans cette philosophie de dire qu'à un moment donné, même si l'étudiant se retrouve seul à la maison, il faut qu'il y ait un prof quelque part. Parce que demander aux étudiants de travailler en autonomie, c'est quand même le truc qu'on demande le plus aux gens sauf qu'on se soucie peu de savoir s'ils savent le faire. C'est une compétence à part entière de travailler seul, c'est pas inné. Moi je m'inquiète à chaque fois que je propose un travail à distance aux étudiants de m'assurer qu'il y a dans leur entourage quelque chose qui vient valider ce qu'ils font, qui vient corriger. Dans ce cas-ci, l'assistant vocal, pour moi, en effet c'est intéressant quand on touche à quelque chose qui est vocal, c'est-à-dire à la prononciation, dans ce cas-ci.

Et donc ça serait utile au niveau de quelle compétence selon vous ? Seulement pour la prononciation ou cela peut être aussi pour autre chose ?

Ici, le travail c'est de produire du son et que l'assistant vocal puisse valider si la prononciation est correcte ou pas, proposer la correction, répéter le mot, etc. Moi, je n'utilise pas d'assistant vocal dans d'autres contextes. J'aurais tendance, très rapidement, à me dire quelle serait la plus-value. La plus-value sur base d'autres compétences serait peut-être d'avoir des étudiants qui préfèrent qu'on leur explique verbalement, plutôt qu'on leur explique par écrit. Je pense que du coup, l'assistant vocal serait intéressant parce que je crois que c'est plus facile pour certains étudiants de comprendre une élaboration parlée qu'une élaboration écrite. Parce que comprendre un texte, comprendre une élaboration sur papier, c'est parfois très compliqué. Il y a peut-être aussi une dimension très émotionnelle, parce que la voix c'est quand même très émotionnel, c'est peut-être plus humain qu'une feuille de papier qu'il faut lire et où c'est compliqué. Du coup, oui, ça peut être applicable pour moi, à n'importe quelle compétence, mais il faut que ça intervienne à un endroit qui est intéressant. Si c'est juste pour faire "ha il y a un assistant vocal sur tout ce qu'on fait" et qu'en fait cet assistant vocal n'amène pas grand-chose

de plus qu'une autre solution, moins coûteuse, il faut que la plus-value soit vraiment du coup importante pour moi.

Je pense alors peut-être aussi aux étudiants qui ont des besoins spécifiques. Quand on pense à des étudiants qui sont dyslexiques, pour qui le langage écrit est compliqué, c'est sûr que d'avoir le même contenu mais avec un médium qui facilite leur compréhension, là c'est certain que pour certains étudiants, le fait de passer par une vocalisation de ce qu'on leur explique ou de ce qu'ils peuvent lire ou autre, ça peut faciliter bien sûr la compréhension. C'est quand même un périmètre très restreint.

Selon vous, où se positionnent les enseignants et les élèves, justement, en termes de l'adoption d'un assistant vocal ? Enfin, pensez-vous que les enseignants et les étudiants sont prêts à utiliser un assistant vocal ?

Moi, je pense qu'on est à la fois dans la situation, que le covid nous a montré. On est à la fois dans une société extrêmement technologique où il y a des gens qui sont extrêmement à l'aise avec plein de gadgets et plein de technologies, etc. Et puis on a une tranche de la population, qui soient étudiants ou profs d'ailleurs, qui n'en touche pas une. Moi, je reste convaincu que s'il y a une plus-value et si l'utilisation a du sens, ça va couler de source. Ça va être plus instinctif. Si on a l'impression que c'est une couche supplémentaire dans la tâche qu'on doit faire et qu'on doit comprendre le mécanisme, que c'est pas hyper organique, qu'il y a un assistant vocal à un moment donné parce qu'on ne voit pas très bien ce qu'il amène en plus que quelque chose de plus classique, je pense que là on perd du monde. a, moi je dirais que c'est la grande leçon que la COVID nous a apprise en termes de pédagogie, d'éducation, c'est que la charge mentale et la charge cognitive, c'est quand même très limité.

Donc, au plus on ajoute des choses, au plus on demande un effort au cerveau de devoir gérer ces différents éléments, donc de nouveau moi, je n'utilise une technologie que quand elle me permet de faire quelque chose que je ne sais absolument pas faire autrement et où il y a réellement une plus-value par rapport à une manière plus classique de faire. Sinon je ne le fais pas.

Et est-ce que vous avez déjà eu l'occasion de faire l'expérience avec les élèves, d'utiliser un assistant vocal ou une application comme Flowchase ?

Oui, donc là cette année, puisque c'est une phase de test, on avait proposé aux étudiants chez qui on avait détecté un problème de prononciation de pouvoir utiliser cette application. Certains l'ont utilisée. Donc ils l'utilisent chez eux, on l'utilise pas au cours, puisque c'est toujours le principe de classe inversée, c'est-à-dire qu'ils préparent avant de venir et règlent le problème, etc.

Donc il y a une masse de travail à faire avant, avec beaucoup de satisfaction en fait apparemment. Moi j'avais été sur le projet initial, qui s'appelait autrement avant, et je sais que les responsables font beaucoup d'améliorations par rapport à ça. Et donc là on avait réussi à avoir une application assez stable, notamment sur la reconnaissance vocale qui posait des problèmes avant puisqu'en tant que bons profs, on est toujours les pires étudiants, on avait piégé le système. Avec une collègue, on avait passé des heures en allant à un colloque, à voir tous les endroits où nous estimions que la prononciation qu'on donnait n'était pas correcte, mais elle était validée par le système. On a donc fait des retours par rapport à ça en disant "ça manque quand même de précision" etc. Mais ça, ça s'est vraiment amélioré et même les étudiants qui sont des informaticiens, donc du coup la technologie ça leur parle, étaient étonnés aussi de la qualité de retour du système, de la manière dont à la fois les choses sont prononcées, la voix n'est pas métallique, elle n'est pas trop électronique, etc. Et puis la capacité de reconnaissance de ce qu'ils disent et le degré de correction était assez précis et correct.

Qu'est-ce qui compte pour eux ? Puisque le but, comme je disais au départ, c'est d'avoir à un endroit, quand on travaille, quelqu'un ou quelque chose dans ce cas-ci, qui valide et qui fait "c'est bon, tu es sur le bon chemin ou tu n'es pas sur le bon chemin". Et c'est comme ça qu'en fait, on se voit avancer, on se voit progresser.

Alors, vous pensez quels peuvent être les freins à l'idée d'utiliser un assistant vocal dans l'apprentissage d'une langue, que ce soit du point de vue des élèves et des enseignants ?

Alors moi je suis la génération de prof qui, quand j'ai commencé à être enseignant, la technologie arrivait à peu près. Et donc la première des craintes, ça a toujours été que la machine

remplace l'humain. Je crois que ça reste encore un tout petit peu le cas. Pour moi, ça recouvre un manque parfois de la part des enseignants de se positionner et de se dire, « qu'est-ce que moi je peux apporter dans le cadre d'une formation et qu'est-ce qu'une machine ou une technologie peut amener dans le cadre d'une formation ? ». Et en fait l'un et l'autre n'amène pas la même chose et sont assez complémentaires.

Donc ça demande un travail vraiment didactique de comprendre qu'une tâche à un moment donné va être plus intéressante à faire en dehors, avec l'aide d'une technologie, et qu'au cours, quand on a des étudiants en face de nous, il faut privilégier ça, ça, ça et ça. Parce qu'en fait ça c'est impossible de leur faire faire en dehors avec n'importe quelle technologie. Donc je pense qu'il y a cette résistance au changement qui dit "il faut que je remette en question les pratiques, il faut que je redéfinisse ma pratique, que je ne fasse pas comme j'ai l'habitude de faire depuis 20 ans". Donc il y a cette question-là.

La deuxième question, c'est la maîtrise aussi. Moi je suis très content parce que je reçois un produit fini, donc quelque chose qui tient la route, sur lequel je sais travailler. Ce sont des linguistes qui travaillent à la base, donc j'ai confiance sur le développement du contenu, il y a vraiment des profs d'anglais là derrière qui réfléchissent etc. Je me dis de nouveau dans l'utilisation de l'outil, il y a les compétences informatiques. Tout le monde n'est pas à l'aise dans l'utilisation, dans ce cas-ci, d'une application. Moi je sais que des applications j'en utilise tous les jours, donc pour moi c'est un univers tout à fait connu, je ne suis pas choqué par le fait d'utiliser une application, je serais pas choqué par le fait de le faire. Mais je ne suis pas certain que c'est le cas au niveau de tous les enseignants. Aussi, on peut croire que maintenant, en 2021, toutes les personnes qui ont 18 ans savent faire tout sur un ordinateur, mais en fait c'est faux. Il y a des gens qui, par leur passé scolaire ou même familial, n'ont pas été initiés à la gestion de la technologie.

Et donc j'en reviens à cette question de charge cognitive qui, là où on veut faciliter une tâche, en fait on la complique. Parce que ça demande à la personne de réfléchir à plus de choses en même temps, et donc c'est difficile, décourageant et alors du coup on éteint son téléphone et c'est bon je ne fais pas l'exercice.

Alors, il y a aussi toujours la question de l'argent, ça c'est sûr. Qu'est-ce que ça coûte ? Qui paie ? Je vois ici moi j'ai mis là l'application au choix pour les étudiants, parce que ça coûtait quand même 12€, ce qui est un budget. On peut poser la question à l'université, si l'université ne devrait pas intervenir dans le prix. Par exemple, moi, pour faire adopter cette application, j'ai dû quand même faire beaucoup de paperasserie et beaucoup de procédures, parce qu'on doit faire des appels d'offres, des études de marché, etc. Plusieurs fois, je me suis dit "c'est bon je vais laisser tomber". Mais bon, je me suis dit non, ça en vaut la peine et puis voilà, on est allés au bout. Mais je pense que ça peut aussi être un sérieux frein. Comment est-ce que le dialogue se fait entre le terrain plus pédagogique et scolaire, et ces gens qui développent des outils qui doivent gagner leur vie fondamentalement.

Et donc vous avez parlé de disparition des métiers, est-ce que pour vous ça peut être un danger de voir l'assistant vocal remplacer les enseignants ?

Non, parce que de nouveau, je pense que c'est une crainte qu'on a que plus ou moins rationnelle, mais de nouveau ce que la COVID nous a montré, c'est qu'il n'est pas souhaitable, pour personne, qu'il n'y ait plus de salles de cours avec des êtres humains qui interagissent. En fait, la machine ne remplacera jamais ça. Donc moi j'ai pas cette crainte puisque même des étudiants disent voilà que rester derrière l'écran c'est compliqué, tout faire de manière virtuelle c'est compliqué. Ou alors, il faudrait une intelligence artificielle vraiment très poussée, mais qui ne remplacera jamais. Je n'ai pas connaissance d'un produit d'intelligence artificielle qui peut gérer, qui a l'algorithme de la spontanéité. Ce n'est pas possible. Donc moi je ne me tracasse pas trop par rapport à ça. Parce que la capacité de s'adapter à une situation de manière spontanée, ça reste quand même très central dans l'apprentissage et dans l'enseignement. Sauf si on ne fait que des cours magistraux, évidemment, parce qu'on répète à chaque fois la même chose. Mais si on veut s'adapter à la personne qui est en face de nous, il y a plein de choses qui passent par des canaux qui ne sont pas électroniques.

Et alors, est ce que vous pensez que, en utilisant des assistants vocaux, est-ce que ça se fera surtout par des enceintes connectées comme par exemple développées par Google ou Amazon, ou alors justement par d'autres applications vraiment conçues à l'effet d'apprendre une langue ?

Moi je crois qu'il y a comme dans tout, c'est-à-dire qu'on peut partir sur les grandes orientations qu'on peut faire de l'assistance vocale sur la grammaire, de l'assistance vocale sur la prononciation, de l'assistance vocale sur un peu tout ce que les étudiants doivent faire. À un moment donné, il y a aussi toute une série de compétences qui sont vraiment très spécifiques, qui ne s'adressent pas toujours à un très grand public, mais qui sont quand même nécessaires. Et là, malheureusement de nouveau je me dis, le but de Google, Amazon, c'est quand même de faire de l'argent. Croire que Google va venir m'aider sur un dispositif où j'ai 12 étudiants en bloc 3 géographie, où ce serait super d'avoir un assistant vocal sur des trucs que seulement des géographes font, par exemple, dans la pratique de leur métier, je pense pas que ça va avoir énormément de succès de Google, qui ne vont pas financer ce genre de truc.

Donc oui sur les grands chemins oui et c'est très bien parce que voilà ça concerne un peu tout le monde. Après, pour le reste, je pense que ça va être plus compliqué, plus limité en tout cas.

Donc vous pour des applications qui seraient utilisées comme outil pédagogique comme ça vous verriez plutôt ça développé par des start-ups comme Flowchase ou autre ?

Moi ce qui m'intéressait chez Flowchase c'est comme je le disais, c'était initialement l'association d'un spécialiste de la technologie et d'une spécialiste de la langue. Et en fait, je ne vois pas comment est-ce qu'on peut travailler autrement que dans cette collaboration, parce que n'importe quel choix technologique doit être filtré par la personne qui connaît la langue et inversement. C'est la première raison.

La deuxième raison, c'est que moi les dirigeants de chez Flowchase, je peux les appeler quand je veux, j'ai leur numéro de GSM pour leur poser toutes les questions que je veux. Les dirigeants d'Amazon et de Google, ça m'étonnerait qu'ils soient disponibles pour répondre à mes questions sur leur application où je leur dis "est-ce que cet endroit-là vous pensez pas que ça serait mieux de faire comme ça parce qu'en fait mes étudiants ils ont essayé, ils ont pas bien compris". Donc c'est un vrai confort aussi, et c'est un vrai plus de pouvoir collaborer avec ces personnes et d'avoir un accès à ce qu'ils disent. Moi je crois que c'est plus intéressant.

Et puis aussi, il y a le côté peut-être un peu philosophique de dire que c'est une start-up, ça vaut la peine de soutenir, c'est une start-up belge, ça vaut la peine de soutenir, ces gens sont supers,

ils remuent ciel et terre pour que tout se passe bien. Moi je sais que je m'oriente clairement plus vers ce genre de profils.

Et puis au niveau protection des données, c'est toujours plus sûr.

Bien sûr, oui, tout à fait tout à fait, absolument.

Oui, ça vous pensez aussi que ça va représenter un frein à l'adoption, le fait que les assistants vocaux peuvent accéder aux données personnelles ?

Bien sûr, oui. Cette gestion de données personnelles, en effet, ça pose problème dans pas mal de choses donc c'est sûr que si on peut avoir une espèce, pas de contrôle, mais en tout cas d'être très au courant de comment les choses sont faites et organisées, c'est mieux. Là je sais que la responsable de Flowchase a voulu contacter mes étudiants pour leur proposer une extension de leur abonnement. Elle me demande d'abord l'autorisation, elle me montre le message qu'elle renvoie en disant « est-ce que tu valides bien, est-ce que je peux leur dire ça ? Est-ce que je les contacte ou tu veux les contacter toi par un biais tout à fait officiel ? Ou est-ce que je peux les contacter par l'adresse mail qu'ils m'ont renseigné quand ils se sont inscrits ? ». Voilà, c'est le genre d'échange qu'on ne peut avoir que quand on connaît les gens.

Alors, comment est-ce que vous voyez le futur dans l'enseignement des langues ? Vous pensez que les assistants vocaux vont prendre de plus en plus d'ampleur ?

Je pense en effet que c'est une bonne chose que n'importe quelle technologie soit de plus en plus utilisée parce qu'on s'est rendu compte quand même qu'il y a une plus-value. Ça permet de faire des choses qu'on ne peut pas faire autrement. Après comme je disais, ça ne remplacera jamais un être humain, donc je crois que c'est une bonne chose. Je ne suis pas certain que ça va prendre plus de place au sens péjoratif du terme. J'espère que ça prendra plus de place, mais une place justifiée, intéressante, porteuse de sens à l'avenir, oui.

Donc ça peut être un vrai outil pédagogique pour vous ?

Oui, ça c'est sûr et certain. Tout à fait oui.

Et quand vous parlez de plus-value, est-ce que vous avez justement une plus-value vraiment concrète en tête ?

La plus-value concrète pour moi dans le système, dans le dispositif qui m'intéresse ici, c'est de permettre à un étudiant qui travaille seul d'avoir ce qu'on appelle en pédagogie informatique, d'avoir cette présence à distance. C'est dire qu'il y a des étudiants qui ont besoin d'avoir quelqu'un qui leur dit « tu fais ça, tu fais ça puis tu fais ça ». C'est tout à fait normal. Le problème dans les études supérieures, à l'université encore plus, c'est qu'on a un discours très fort qui dit « OK, vous faites ça et vous vous débrouillez parce que ça c'est l'unif et il faut être au top ». Alors il y a des étudiants qui réussissent extrêmement bien parce qu'ils ont probablement été baignés là-dedans depuis le début de leur vie. Et puis il y en a d'autres qui par la force des choses, on ne leur a jamais demandé de gérer quelque chose seul. Donc ça c'est mon questionnement à moi. Donc l'assistant vocal c'est le Petit Jiminy cricket sur l'épaule. En fait, c'est le truc qui va faire qu'on va pouvoir demander à cet assistant vocal d'intervenir à des endroits bien spécifiques pour des fonctions bien spécifiques : corriger, encourager aussi, dire aux étudiants « vas-y continue, c'est bien, c'est super, t'es à la moitié, tu vas y arriver », faire une explication, etc. Donc pour moi la plus-value c'est cette présence à distance. C'est le fait de pouvoir créer un encadrement de chaque individu là où il se trouve sans que le prof doive les appeler, dire « ça va, t'as fait ton truc » parce que c'est ingérable. Pour moi, c'est ça la plus-value.

Et je dirais aussi, mais j'ai moins de recul par rapport à ça, au niveau des étudiants alors soit à besoins spécifiques soit à profil plus spécifique. Moi je sais, par exemple, que si je pouvais choisir entre une explication, comme je disais tout à l'heure et un document à lire, qui m'explique un truc et quelqu'un qui m'explique un truc vocalement, je vais clairement aller vers quelqu'un qui m'explique un truc. Donc la plus-value serait là peut-être de proposer une alternative à quelque chose de très sacré comme le texte écrit. Moi j'ai fait pour les étudiants, pour l'année prochaine, un descriptif de cours écrit, puis un descriptif de cours où je me suis enregistré et j'explique. Comme ça, ils choisissent ce qu'ils préfèrent comme *medium*.

Et est-ce que vous pensez que l'utilisation d'une application comme Flowchase ou d'un autre assistant vocal qui aiderait à l'apprentissage d'une langue doit être utilisée en complément

d'un enseignement universitaire par exemple ? Ou ça pourrait très bien être utile pour une personne qui décide comme ça d'apprendre l'anglais de manière autodidacte ?

Alors ça, on touche à la didactique des langues étrangères. Une langue, c'est fait pour être utilisée. On se rend compte que dans beaucoup de cours de langue, ce qu'on explique aux gens, c'est comment la langue fonctionne : la grammaire, le vocabulaire, etc. Comprendre comment une langue fonctionne et est utilisée, c'est deux choses différentes. Donc, comme j'explique toujours aux étudiants : si on vous avait appris à nager en faisant un cours magistral au bord de la piscine, en vous expliquant comment on met les bras et les jambes, une fois qu'on vous pousse dans l'eau, ça ne veut pas dire que vous êtes capable de nager. Du coup ici, il y a une vraie complémentarité entre : je développe les connaissances dont j'ai besoin et j'essaie d'utiliser les différentes briques que je vais devoir mettre ensemble. Et construire un mur, c'est autre chose. Donc pour moi c'est vraiment de la complémentarité. Je ne verrais pas l'intérêt de Flowchase pour quelqu'un qui se dit « ha, j'améliorerais bien ma prononciation ». C'est étrange parce qu'améliorer sa prononciation... pourquoi faire ? Dans quel cadre ?

Moi je sais que j'utilise Flowchase en expliquant aux étudiants qu'ils vont étudier des trucs très techniques, par exemple, les *focus words*. Donc les *focus words* c'est le fait de mettre, à un moment donné, plus d'intonation sur un mot pour le faire ressortir dans la phrase. Ce que Flowchase explique, c'est comment on fait. Après si au cours il n'y a pas un exercice ou on leur demande de faire une intervention orale où ils doivent utiliser notamment des *focus words* et que ça leur serve dans leur communication, ça n'a pas beaucoup de sens en fait. C'est comme étudier de la grammaire en faisant des exercices. C'est pas ça une langue. Une langue, c'est au moment où je vais communiquer un truc, je vais choisir une forme grammaticale qui va me permettre de traduire une signification bien précise. Donc ça pour moi c'est hyper complémentaire. Du coup, ce que j'appelle les outils de la langue, ça peut se faire via Flowchase ou autre. Mais à un moment donné, il faut investir dans un travail, plus de compétences en fait. Donc, complémentarité, oui, ça c'est sûr.

Une autre question qui me vient en tête, c'est plus en rapport avec les élèves et l'enseignant. Vous pensez vous que l'assistant vocal va plus assister l'étudiant à apprendre une langue et moins à aider l'enseignant à changer sa méthode ?

Ouais, moi je pense. Je pense qu'on a recours à l'assistance vocale comme on aurait recours à des assistants. C'est quand on veut autonomiser la personne et qu'elle ne doit pas systématiquement revenir vers nous pour avoir un feedback. C'est au même prix que de la correction automatique. C'est le même principe.

Pour moi ça serait vraiment parce que ça permet de déléguer une tâche en toute confiance. Du coup ça permet à chacun d'avancer à son rythme. Après cette première étape de travail, on avance sur d'autres choses. Ou si un individu a besoin d'aller un peu plus loin dans le travail, il peut revenir vers le prof. Je trouve que c'est un bon filtre. En fait, c'est une bonne première étape de travail.

Ok, merci beaucoup, je ne sais pas si vous avez aussi quelque chose à ajouter.

Non, je ne pense pas. J'avais lu vos questions, c'était clair donc.

Ok merci beaucoup en tout cas.

Merci à vous et bon courage.

Merci bonne journée. Au revoir.

2. Interview Robin Guérit

Bonjour, vous m'entendez bien ?

Bonjour, oui.

Enchanté.

Enchanté, merci d'avoir accepté de participer à cette interview.

Y a pas de souci ! C'est un mémoire que vous rendez quand ?

En août, début août.

Ok. Et vous êtes deux à travailler dessus sur le même master en fait ?

Oui, c'est ça. On est en master en sciences de gestion à Louvain-la-Neuve.

Ok ! Je viens de l'UCL aussi donc je vois bien. Et comment vous voyez la réunion ?

En fait, on va peut-être d'abord commencer par se présenter et dire le but de notre mémoire. Donc on est Alice et Pauline et on est toutes les deux en sciences de gestion à Louvain-La-Neuve. Et le but de notre mémoire, c'est d'analyser comment l'assistance vocale va apporter une plus-value dans 3 cas d'usage différents. Et donc on a choisi de travailler sur les cas d'usage d'aide à la conduite, l'apprentissage d'une langue et les secours à distance. Donc on vous a contacté pour le deuxième cas d'usage, donc qui est l'apprentissage d'une langue d'une langue étrangère.

Ok très bien. Je peux me présenter rapidement aussi. Et puis on peut regarder aux questions au niveau de l'heure. Je voulais juste préciser qu'en fait j'ai une réunion à 16h30, mais on peut discuter ensemble et si vous avez des questions plus précises, envoyez-les-moi, j'y répondrai demain matin. Donc on fait ça et si vous avez des questions pour lesquelles il faut développer

un peu plus, je peux mettre quelques références aussi si j'en ai, je le ferai. Enfin peut-être pas 25 questions non plus mais voilà vous avez compris.

Oui, du coup moi je m'appelle Robin, j'ai cofondé Flowchase et principalement nous, ce qu'on fait, c'est qu'on booste la prononciation des étudiants grâce à une technologie vocale, du coup ou pas via Siri, Alexa ou autre mais en fait nous en fait on développe en interne une technologie qui permet d'analyser la voix, détecter les erreurs et donner des feedbacks personnalisés. Et en fait, ce feedback, il est vraiment travaillé avec des linguistes, des professeurs et avec des ingénieurs évidemment pour vraiment trouver les meilleurs conseils à donner aux étudiants au moment le plus opportun. Et donc ce qu'on fait, pour être un peu plus précis parce que c'est un peu le sujet, c'est comment ça fonctionne vraiment.

En fait, chez nous, on a tout un programme d'apprentissage, ce qu'un étudiant va faire. On a des activités qui introduisent un peu au concept de prononciation, qui sont très rapides, avec un quiz à la fin pour essayer d'un peu mémoriser. Donc ça, c'est très rapide. Puis on a des exercices d'écoute, parce que travailler la prononciation, ça va de pair à avoir une bonne oreille, détecter les sons, l'intonation etc. Et puis on a des exercices de *speaking*, où là la technologie va vraiment analyser la voix de ce que les gens disent. Et ça c'est nécessaire pour apprendre aux étudiants à vraiment maîtriser, avoir vraiment cette compétence au niveau de la prononciation. C'est pas juste dire certains mots parce que j'ai envie un jour de dire le mot "hiérarchie" et je vais aller sur Google Translate. C'est vraiment de les accompagner dans un programme avec des objectifs précis en fait. Donc ça c'est un peu le contexte.

Et du coup, plus précis, parce que c'est toujours la question ; les « *speaking exercises* » chez nous, ils disent des mots, des expressions ou des phrases entières. En fait, on vise un problème de prononciation en particulier. Donc si vous dites une phrase ou un mot, on ne va pas vous dire « Ah oui, en fait là vous avez 26 erreurs, du coup on va aller revoir chaque erreur, voici 26 conseils ». Non c'est pas notre but, notre but c'est pas de décourager les gens. Notre but c'est vraiment de les motiver en disant « si vous améliorez ça, vous êtes compris ». Et donc dans le programme que je vous ai dit, on a vraiment envie de donner des objectifs par thématique de prononciation. Donc on va d'abord travailler de l'intonation, l'accentuation puis travailler certaines voyelles et consonnes qui sont très importantes à maîtriser pour être compris. Et donc

on fait déjà ce travail de sélection et de réflexion pour les étudiants, pour ne pas les noyer dans plein de trucs qu'il faudrait apprendre pour avoir une bonne prononciation.

Notre but, c'est vraiment de travailler sur les points importants où on peut vraiment avoir beaucoup plus de marge de progression, pour être compris et mieux parler anglais. Voilà, maintenant je peux répondre à vos questions évidemment sur le concept, la réflexion, les retours qu'on a pu avoir etc., mais voilà c'est ce que je voulais dire.

Ok super. Du coup comme on en avait discuté un peu via LinkedIn, vous aviez voulu utiliser un assistant vocal et vous vous êtes rendus compte que c'était pas la meilleure méthode d'apprentissage ? Et ensuite vous êtes dirigés vers la technologie que vous utilisez maintenant ?

C'est ça, en fait au tout début, donc on est un projet universitaire, on est un projet Spin-Off. Donc ça veut dire qu'on a développé une technologie à l'université et qu'après ça, on a vraiment voulu lancer l'entreprise. Et dans ce programme Spin-Off, on a testé beaucoup de choses et on a travaillé avec beaucoup de professeurs. Et donc c'est pas nous qui, personnellement, avons testé les Alexa et les Google Translate, c'est plutôt des professeurs et le retour des professeurs et des étudiants avec qui ils ont cours. Pourquoi ? Parce que c'était la question, c'était quoi notre concurrence, nos substituts et du coup c'est pour ça qu'on avait déjà testé ça avant. Et c'est aussi la raison pour laquelle on croit que notre technologie peut apporter beaucoup plus que, par exemple, les assistants vocaux comme Alexa.

Alors pour être un peu plus concret, au tout départ on avait un prototype de BD interactives, et on comparait ça aux répétiteurs qui existent à l'échelle. Je sais pas si vous voyez c'est quoi le répétiteur, c'est dans les cours d'Anglais qu'y a à l'UCL ? En gros ça veut dire que nous au départ, on avait un prototype pour raconter une histoire et les gens devaient prononcer certains dialogues, et nous, on les corrigeait plus ou moins sur ce qu'ils avaient dit. Et comparé à ça, on avait pris le mot commun qui était dans un répétiteur à l'UCL (c'était y a longtemps, je dis ça c'était il y a 3 ans) et on avait demandé aux étudiants « qu'est-ce que vous préférez faire ? Est-ce que vous préférez le répétiteur à l'UCL, dans Moodle, où vous avez juste certains mots et vous les répétez, puis vous comparez, mais où il n'y a pas vraiment une/de(?) technologie derrière ? Ou est-ce que vous préférez utiliser notre prototype avec une petite histoire, un petit

dialogue et où vous recevez un peu plus de conseils sur des objectifs précis ? ». Évidemment, notre prototype a plu un peu plus à ce niveau-là, il n'était vraiment pas parfait et en fait on a beaucoup changé depuis ça puisqu'on ne fait plus d'histoire. Mais c'est là qu'on a vu que les étudiants étaient intéressés à avoir un feedback un peu plus précis, d'avoir des conseils, et pas juste un pourcentage ou même juste pouvoir se comparer comme dans le répétiteur qui n'était pas assez intéressant pour les étudiants.

Donc ça c'est le genre de tests qu'on a fait. Évidemment, depuis 3 ans on a fait beaucoup plus de tests que ça parce qu'on a voulu développer notre produit en se disant qu'on est en complément de la tâche du professeur. Et du coup, une des questions qui s'est posée pendant le développement du projet c'est « est-ce qu'on remplace le professeur, ou est-ce qu'on accompagne le professeur et on est complémentaire à lui ? ». Et évidemment nous, dans notre optique, c'est d'être totalement complémentaires. On croit à fond au futur de l'éducation avec des professeurs et pas juste avec des robots. Et du coup, on a vraiment créé un outil qui vient compléter le rôle du professeur, et *paf* au niveau de la prononciation, ils sont débordés parce que corriger la prononciation devant 20-40 élèves devant eux en 1h, c'est impossible. Et même dans un cours, s'il les voit 2h par semaine par exemple, chaque étudiant n'a pas forcément l'opportunité de parler. Du coup, on a vu vraiment l'opportunité de développer notre outil en complément avec des objectifs précis.

Donc première chose que je voulais dire, c'est qu'on a vu que les étudiants voulaient du feedback beaucoup plus personnalisé, beaucoup plus orienté vers des objectifs à atteindre. C'est des mini-objectifs, mais c'est beaucoup plus valorisant pour se sentir progresser. Et la deuxième chose, c'était qu'on voyait qu'on avait beaucoup plus de chance que le produit marche si on était en complément d'une formation. C'est-à-dire que l'étudiant fait de la grammaire, du vocabulaire et est déjà dans un cursus, mais va utiliser un outil de manière individuelle chez lui pour travailler uniquement cette prononciation. Et puis va revenir en cours pour pas faire de la prononciation tout seul mais pour continuer à développer cette prononciation avec les autres, avec des activités de groupe que nous, on donne aussi avec un manuel au professeur. C'est pour ça qu'on est un peu hybride, on a cette app à la maison avec tout ce qu'on a appris, et tous nos exercices. Et on a ce manuel pour les profs où ils sélectionnent simplement des activités de groupe, comme ça ils ont plus facile à organiser l'interactivité en classe.

Voilà, désolé, c'était une longue réponse. Mais je réfléchis au test qu'on a fait concrètement et oui, évidemment Google Translate en fait, c'est très bien parce qu'on peut s'enregistrer, on peut avoir un feedback, mais c'est plutôt comme un pansement en fait. Dans le sens où je vais aller chercher un mot que j'ai besoin pour ma présentation et je vais écouter comment on le dit, je vais le répéter une fois, mais du coup après on le perd, on l'oublie, on l'a fait une fois parce que le lendemain je dois pouvoir le dire mais ça ne rentre pas du tout dans la courbe d'apprentissage dont on a besoin pour vraiment après avoir une nouvelle compétence. Que ce soit Google Translate ou même Alexa. Alexa il y a plusieurs tests, je peux essayer de retrouver des références si vous voulez.

Mais concernant vraiment les agents conversationnels, Alexa et tout ça, il y a des études qui ont été faites par des professeurs d'anglais qui ont essayé d'intégrer ces outils-là dans des cours pour travailler la prononciation ou l'oral. Et du coup ils leur ont demandé de faire des dialogues de conversation avec la machine. Les professeurs disaient « demandez la météo, de mettre telle musique ou d'aller chercher quelque chose sur internet ». Et sur base de ces exercices de conversation, si Alexa ne comprend pas ce que vous dites, ça veut dire que votre prononciation n'est pas bonne. En gros c'est ce qu'ils ont fait, mais ils ont vu évidemment des limitations. Déjà, il faut avoir Alexa et pas juste un téléphone. Deuxièmement, je ne suis pas exhaustif, mais y a pas vraiment de feedback, du coup souvent la réponse c'est « je n'ai pas vraiment compris votre requête », parce que les prononciations ne sont pas correctes, ou alors il comprend des requêtes mais qui sont erronés donc il ne fait pas exactement ce que la personne lui dit. Mais du coup on n'est vraiment pas dans quelque chose de réfléchi et pédagogique. L'inverse de tout ce que nous on veut faire. Voilà.

Oui, donc selon vous, les assistants vocaux des GAFAs, ou d'autres grandes entreprises, c'est pas les plus appropriés pour assister un élève à apprendre une langue ?

Nous, on réfléchit par objectifs d'apprentissage. Et comme on ne peut pas mesurer les objectifs d'apprentissage via Alexa, c'est très compliqué d'évaluer. Ça, c'est vraiment le premier constat. Le deuxième constat, nous on crée un outil pédagogique qui a été réfléchi avec des professeurs. Aujourd'hui, on a des agents conversationnels qui ne sont pas réfléchis pour le contexte d'apprentissage. Donc, actuellement je ne trouve pas qu'on puisse résoudre ces objectifs d'apprentissage avec les outils. Après, plus tard je ne sais pas, ça pourrait être intéressant de

continuer peut-être à demander à des étudiants de faire des activités un peu rigolotes sur leur prononciation avec Alexa, mais ça ne peut être que divertissant et pas pédagogique au final. Je pense que ça peut être marrant et engageant pour dire à un étudiant « essaie de faire un peu une autocritique de ce que tu as pu avoir comme conversation ». Si ça tombe elle est nulle part ou si ça tombe, elle était sympa. Et ça, ça marche autant avec les bots textuels en anglais qu'avec des agents conversationnels vocaux. Donc ça existe évidemment. Il y a des chatbots, maintenant, on peut apprendre l'anglais à l'écrit, mais évidemment, ça part un peu dans tous les sens aussi. On n'a pas l'objectif d'apprentissage, c'est engageant, ça peut être divertissant, là-dessus je ne suis pas contre. Mais, comme c'est pas cadré et comme on ne sait pas vraiment évaluer où on va, au niveau pédagogique ça n'a pas vraiment d'apport. Tout dépend de ce qu'on recherche en fait.

Donc pour vous, pour qu'une technologie soit vraiment utilisée comme un outil pédagogique, il y a vraiment ce besoin de feedback, de retour du système pour l'étudiant et aussi de complémentarité avec un professeur ? C'est ça qui est le plus important ?

Oui c'est ça. Après nous, on vient de l'université, donc on s'est inspiré autant de la technologie de papier que de la technologie vocale, et qu'est-ce qu'on pourrait faire au mieux pour amener des conseils. Et aussi, on s'est basé sur toute la recherche en linguistique appliquée, où il y a un champ de recherche là-dessus qui est 1) comment les outils numériques peuvent aider à travailler la prononciation, et 2) quels sont les problèmes de prononciation qui sont à travailler en priorité pour tous les étudiants. C'est vraiment quels sont les problèmes à régler en premier. Donc ça c'est vraiment toute cette recherche-là.

Donc évidemment nous, oui on est tout à fait convaincus qu'il faut du feedback personnalisé et pas juste de la comparaison. Parce qu'un étudiant qui n'arriverait pas entendre les différences ne va pas pouvoir s'auto-corriger simplement en écoutant un modèle en fait. Donc c'est pour ça qu'on croit avoir ce truc supplémentaire qui permet aux étudiants de savoir s'ils doivent corriger ça de manière forte ou s'ils doivent corriger quelque chose dans leur bouche pour dire tel son *I*. Et en fait, nous, on a une batterie de conseils, donc ce n'est pas juste un conseil. Notre but c'est vraiment de trouver les meilleurs conseils pour chaque thématique. Ce qu'on fait c'est qu'on a des dizaines d'alternatives de conseils, qu'on teste dès qu'il y a une erreur. Et maintenant on arrive à identifier quel(s) conseil(s) on préfère mettre en avant aux étudiants pour qu'ils

s'améliorent plus rapidement. Il faut voir ça vraiment comme un arbre de réponses en fonction des erreurs des étudiants. Donc en fonction de l'erreur précise d'un étudiant.

Je vais prendre un exemple qui est le mot « *comfortable* ». Si on travaille l'accentuation dans un mot et qu'on dit « *comfor-TA-ble* », évidemment nous on va dire « tu as accentué là ». Il voit sur une visualisation, donc c'est super clair. Et on leur dit « essaye de faire la première voyelle un peu plus forte, un peu plus longue ou regarde une vidéo où il y a quelqu'un qui le fait ». On essaie vraiment de donner des cartes de feedback vraiment chouettes, engageantes, mais qui sont efficaces et qui vont plus que juste écouter un son et d'écouter le sien.

Et est-ce que c'est déjà beaucoup adopté, ce genre de pratiques auprès des étudiants ?

Auprès des étudiants oui et non. En fait, il y a des applications qui commencent à faire ce genre de choses, donc qui connaissent une technologie pour corriger les étudiants. Toutes les applications ultra connues aujourd'hui ne font pas des conseils personnalisés. Ce qu'elles font souvent c'est un pourcentage. Donc sur un mot, une phrase ils disent « vous avez bien prononcé la phrase à 68% », par exemple. Mais tu ne sais pas exactement ce que tu dois améliorer en priorité. C'est intéressant parce qu'on peut toujours se comparer, mais c'est assez limité.

Puis on a des concurrents directs qui font ça dans une autre app qui s'appelle Elsa Speak, qui font plus ou moins la même chose que nous. C'est vraiment notre plus grand concurrent. Eux sont en B2C, nous on est en B2B. Eux sont très connus en Amérique et au Vietnam par exemple. Ça fait 5 ans qu'ils existent, et ils font plus ou moins la même chose que nous, ils corrigent en identifiant les erreurs dans les mots et dans les phrases. Évidemment, je suis totalement biaisé. Je vais dire qu'on fait des meilleurs conseils qu'eux. Mais, mais ils font bien ça, ils ont développé une bonne technologie pour identifier les erreurs, ça c'est vrai.

En général, après les autres plateformes de langue ne vont pas encore vers ce genre de technologie pour la prononciation. C'est pas assez précis parce qu'ils font souvent en grammaire, vocabulaire, *listening*, expression orale, prononciation. Et du coup ils font tellement de choses, ils font de tout, parfois bien parfois pas très bien, mais ils font de tout donc ils vont pas aussi précisément que nous pour travailler l'oral. Nous on fait que de l'oral.

Donc vous c'est uniquement pour la compétence prononciation ?

C'est ça. Et pour vraiment donner un outil. On commence donc pour les étudiants, on essaie de faire vraiment un produit qui est engageant, sexy, fun. Donc on retravaille vraiment ça, que ce soit dans les textes qu'on écrit, que ce soit même quand ils réussissent ou quand ils ratent. On va avoir une fonctionnalité en septembre qui donne un gif avec un truc d'encouragement et tout. Notre but, c'est vraiment de faire du feedback émotionnel aussi pour qu'ils soient motivés à apprendre, que ce soit fun et aussi un feedback efficace.

Et évidemment, c'est un produit complémentaire avec les professeurs, et ça c'est plutôt une stratégie de notre côté où on veut vraiment que ce soit un outil qui est validé pédagogiquement par des institutions qui sont renommées, comme les universités. C'est pour ça qu'on fait ça et qu'on accueille et qu'on accompagne ces professeurs à utiliser l'outil et à le prescrire. Mais ça c'est vraiment juste une stratégie. Il n'y a pas vraiment de papier scientifique qui va vous dire que c'est la meilleure solution. Nous, on est convaincus, sur base de la recherche qu'on connaît, sur base des retours qu'on a, mais évidemment, il y a personne qui le dira de manière totalement objective. Peut-être que dans 2 ans je pourrais vous le dire, quand on pourrait avoir assez de recul et demander des études externes pour évaluer l'impact de l'outil mais actuellement je peux pas vous le donner.

Ok. Une question un peu technique. Donc vous, votre technologie, elle utilise pas du tout le même système qu'un assistant vocal ? Qui peut quand même identifier les erreurs de la reconnaissance vocale.

C'est plus ou moins les mêmes principes et en même temps c'est pas les mêmes principes. Disons qu'on prend les premières étapes de la reconnaissance vocale. Je veux pas être trop technique, mais on peut reconnaître les phonèmes qui sont les sons de la langue, mais nous on ne va pas forcer la reconnaissance comme pourrait le faire Siri. Parce que, par exemple, si vous dites des choses qui sont proches de ce que vous voulez dire, Siri va comprendre en fonction du contexte ce que vous voulez dire. Même si parfois c'est pas forcément ce que vous avez dit. Donc en fonction des contextes, ils vont biaiser la manière dont ils vont comprendre, parce que eux, ils veulent comprendre le plus rapidement ce que vous avez dit. Ils ne sont pas là pour vous apprendre.

Nous il faut vraiment voir ça comme : on prend les premières étapes de la reconnaissance vocale, on voit au niveau des phonèmes, et après nous on fait vraiment tout un travail fonction de ce qu'on a identifié comme erreur et en fonction de ce qu'on voulait que l'étudiant dise dans cette phrase. On développe quelque chose qui nous permet de pouvoir extraire cette erreur et sur base de cette erreur, leur donner le meilleur conseil. Et ça c'est totalement différent de la manière dont eux ils travaillent, parce qu'on n'a pas du tout le même objectif final. Il y a des objectifs de reconnaissance, il y a des technologies qui ont des objectifs de text-to-speech, donc de pouvoir traduire directement et de tout mettre en texte, etc. Nous, c'est pas notre but. Nous, on vous prend le son brut, on le découpe en gros, et en fonction de ce qu'on entend comme son, on va connecter ça en fonction de ce que l'étudiant devait dire. Et après il y a toute la partie de conseils et feedbacks derrière. Donc oui et non.

Et alors quels sont, selon vous, les freins à avoir ce genre de technologies ? Que ce soit même pour les étudiants et pour les professeurs ? Enfin, je suppose qu'il y a des professeurs qui n'ont pas forcément envie d'avoir ça comme outil complémentaire ?

Au niveau des freins, c'est compliqué de dire par rapport à la technologie. Je réfléchis, mais je vais dire que, en général, dès qu'un professeur voit notre outil, il dit « Ah, ça existe ça, c'est trop bien ». Ils n'ont rien à faire de leur côté, parce qu'on crée le contenu pour eux. Il est réfléchi au niveau pédagogique, il est testé pour ça. On les assiste pour accompagner dans leur formation, on leur donne les clés pour engager les étudiants, donner un tout nouvel outil fun sur le mobile pour travailler de manière individuelle. C'est un outil un peu rêvé pour les professeurs.

Les freins par rapport à la technologie, évidemment, il y a toujours « est-ce que ça va marcher pour tout ? ». Évidemment, ça c'est des choses auxquelles on ne sait pas répondre « oui, ça marche pour tout à 100% ». On a toujours un pourcentage d'erreurs. Il y a toujours un potentiel pourcentage que la technologie n'a pas fonctionné sur tel enregistrement. Donc oui ça peut être un frein. C'est évidemment un truc sur lequel on travaille maintenant mais qu'on travaillera toujours dessus. Donc ça, ça peut être un frein. Et pour ça qu'est-ce qu'on met en place pour résoudre ça, c'est vraiment de faire des tests à grande échelle avec plein d'enregistrements qui viennent de partout, pour tester notre technologie et améliorer le pourcentage de correction. En tout cas, la maturité de la technologie. Et du coup ça c'est ce qu'on fait en interne.

Sinon, je réfléchis à d'autres freins. Mais c'est ça, c'est pas lié à la technologie directement. C'est plutôt lié à nous, dans notre phase de développement. Aujourd'hui, on travaille sur du contenu pré-écrit, donc on fait des programmes personnalisés en fonction des niveaux et en fonction des thématiques. Donc ici, on travaille sur de l'anglais général, et là on va avoir des programmes en business English justement de là où vous êtes. Donc on va customiser ces programmes-là, mais c'est toujours sur base de contenus pré-écrits, les mots les plus importants pour tel secteur à maîtriser, etc. Et donc on écrit des phrases, on écrit des mots, mais évidemment, c'est pour assister les étudiants sur le vocabulaire qui est utilisé en général pour leur secteur.

Si demain vous voulez travailler à votre présentation et dire une phrase bien précise, aujourd'hui, on n'a pas encore la technologie pour pouvoir le faire. On l'aura, mais on l'a pas encore. Donc ça, c'est vraiment l'un des freins qu'on a plutôt du côté étudiant, mais c'est pas un frein par rapport à ce qu'on veut faire. C'est juste un frein parce que ce serait vraiment génial d'avoir une technologie qui fonctionne en fonction de ce que je dis maintenant, ou en fonction de ce que j'ai envie de dire pour préparer ma présentation. C'est vraiment ça.

Je réfléchis, mais qu'est-ce qui pourrait être un frein à l'utilisation de cette technologie ? Ce qui pourrait être un frein en fonction de la technologie, c'est au niveau des accents. Mais c'est vraiment un truc qu'on n'a pas encore eu comme erreur. On n'a pas vraiment identifié ça maintenant. Mais potentiellement on pourrait avoir des problèmes en fonction des différents accents dans le monde entier. Et on ne parle pas des accents natifs, on parle d'accent d'un japonais qui parle anglais et d'un turc qui parle anglais, d'un espagnol qui parle anglais. Et ça, ce qu'on fait, c'est qu'on travaille des alternatives. Parfois on accepte ou on refuse plus d'erreurs en fonction d'où ils viennent. Donc ça, c'est encore quelque chose qu'on travaille. Ça peut être une limitation, parce qu'il faut vraiment entraîner la technologie. En fait, il faut des données. La solution là pour ça, c'est des données. Il faut annoter, valider en fonction d'où les gens viennent comme langue maternelle, qu'est-ce qu'on peut faire pour améliorer ça ? Donc ça, ça pourrait être des freins aussi de la technologie. Mais on est conscients qu'on a pas ce problème là pour le moment.

Et peut-être au niveau des limites dans votre technologie et aussi les limites des assistants vocaux peut-être la raison pour laquelle vous ne vous êtes pas tournés là-dessus ?

Comme je l'avais dit tantôt, pour moi, c'est la seule justification que je vois. C'est que ça peut être dans un cadre engageant et divertissant mais pas cadré pédagogiquement, simplement. On progresse que si on évalue correctement. L'évaluation est très importante, que ce soit pour un professeur ou simplement l'étudiant lui-même. Même si on imagine avec les agents conversationnels, pour corriger un minimum ou dire « j'ai pas très bien compris ce mot là, tu peux répéter ? ». C'est compliqué aujourd'hui de se dire que l'étudiant va progresser parce que s'il le fait une fois tous les 3 mois,... C'est plus divertissant et engageant à se dire « Ah oui, ça me motive à apprendre l'anglais » que « ah oui j'ai vraiment appris quelque chose ». C'est vraiment ça ma plus grande peur par rapport à ça. Après, tout dépend des logiciels qui vont être annexés et qui vont être secondaires à cette technologie-là. Évidemment il y a sûrement des projets par ci par là qui existent autour de ça, parce qu'ils font de la recherche dans tous les sens. Mais voilà, moi je pense que c'est vraiment ça la grande limitation pour que l'étudiant, voire même les professeurs, voient la plus-value de la technologie.

Et comment envisagez-vous le futur des technologies vocales dans l'apprentissage d'une langue ? Est-ce que ça va être de plus en plus répandu et de mieux en mieux performé ?

Oui, je pense que clairement comme on voit ici au niveau de l'apprentissage au niveau du texte, comment on peut corriger déjà des textes en anglais, évaluer le niveau du texte, trouver des erreurs grammaticales, de sens, d'expressions, etc. Je pense qu'on pourrait aller vers ça plus tard avec le vocal. C'est juste que le vocal est plus compliqué à développer que le texte. Enfin, ça dépend, c'est pas la même techno. Mais je pense que oui, ce sera vraiment quelque chose de répandu et efficace. Il faut juste voir ce qu'on veut en faire en fait.

Après, oui, je pense qu'à ce niveau-là, il n'y a pas trop de limites. Après tout dépend si on parle d'apprentissage. Si on parle de l'apprentissage, oui je pense qu'on aura encore plus de concurrents après, mais je pense qu'on sera 50 dans l'équipe et que du coup, on sera bien. C'est comme ça que je le vois.

Sinon ça existe aussi un produit qui ressemble à nous, mais pour la lecture chez les enfants. C'est une entreprise qui s'appelle Lalilo. Donc LALILO, qui font la même chose que nous mais depuis 4-5 ans pour assister la lecture des enfants en français et en anglais. Et du coup, je pense que oui, on va continuer à avoir des outils comme cela parce que quand ils sont bien réfléchis

pour l'apprentissage, et que les professeurs et les étudiants voient la plus-value à la fin, oui, clairement c'est une opportunité énorme. Évidemment, la peur c'est toujours que les robots nous remplacent, mais en fait ils ne réfléchissent pas eux-mêmes, c'est nous qui réfléchissons pour eux. Quand même, c'est nous qui les créons. Donc s'ils sont vraiment créés dans un cadre où on sait que ça va être pédagogique et qu'on va amener vraiment quelque chose de positif et d'efficace aux étudiants, évidemment ce sera toujours dans un bon sens.

J'imagine que dans les prochaines années on verrait bien, sur Club House, je sais pas si vous voyez Club House ce que c'est ? C'est un peu une room mais que avec de l'audio. Évidemment peut-être que plus tard, dans 2-3 ans, on verra un prof d'anglais, qui est en fait un robot vocal, qui parle avec des gens pour leur apprendre l'anglais juste en faisant une discussion comme ça. Je pense que ce serait possible mais après, à quoi ça sert ? Je ne sais pas s'il remplacera vraiment le prof et je ne sais pas si, en termes techniques, ça peut être intéressant. Mais en termes pédagogique c'est pas vraiment intéressant parce que si on apprend une langue c'est pour se connecter aux gens. Donc il y a des professeurs pour ça et j'y crois vraiment très fort.

Au niveau des assistants vocaux, voilà. Dernière chose et après je vais devoir y aller. Le dernier plan technique que je vois vraiment efficace, c'est la personnalisation. Là on va pouvoir détecter des erreurs de chaque étudiant, vraiment aller personnaliser les programmes ou les choses à apprendre, et évidemment là la technologie peut être vraiment très efficace. Mais toujours moi, je le vois dans une vision d'assister les professeurs pour dire « dans ces 55 élèves, regardez la majorité de la difficulté c'est ça. Si vous voulez, on va leur donner un programme tous de manière individuelle qui est différent, ou soit vous pouvez, dans votre classe maintenant, connaître les difficultés majeures de votre classe et agir dessus ». Cette technologie peut vraiment nous aider à améliorer la manière dont on va mener, changer les objectifs d'apprentissage. Et je reviens toujours à ça. Parce que les objectifs d'apprentissage, c'est la manière dont on évalue, c'est la manière dont on voit le progrès. La technologie plus tard pourra vraiment être cette performance là-dessus. Dans l'individualisation.

Si vous avez d'autres questions, n'hésitez pas à me les envoyer. Je peux vous envoyer quelques références, il n'y a pas de problème. Voilà je sais pas si vous avez d'autres questions qui sont vraiment sur une autre thématique ou sur une autre dimension du projet mais voilà.

Je pense qu'on a fait le tour. Et puis vous avez été complet.

Et vous enregistrez la session ou pas ?

Oui, on a enregistré, mais ça peut être confidentiel.

Ah non y' a pas de problème. Aucun problème.

Ok super.

Mais dites-moi, n'hésitez pas à m'envoyer un truc.

Super merci beaucoup. Et si vous avez des personnes de contact qui pourraient être intéressantes peut-être pour pouvoir répondre à nos questions, parce qu'on aime bien avoir plusieurs points de vue, et pour notre échantillon c'est important.

Moi je vous dirais d'office de contacter Lalilo, mais je ne sais pas si vous auriez une réponse. Parce que c'est vraiment ça, il y en a pas beaucoup qui font, sinon c'est des grosses entreprises où vous n'aurez pas de réponse. Vous pouvez essayer. Je vais réfléchir, je verrai demain si j'ai une idée de quelqu'un qui peut en parler.

Par exemple, on a interviewé François-Xavier Fievez, qui a utilisé votre technologie dans le cadre de son cours. Et donc il nous a bien répondu aussi, c'était intéressant d'avoir son point de vue en tant que professeur aussi.

Ah vous l'avez interviewé ?

Oui.

Ah, c'est intéressant ça. Je suis intéressé par ce qu'il a dit en fait.

Il était positif. On peut vous envoyer ça aussi.

Je ne sais pas si ça reste confidentiel, mais je suis intéressé. Je ne les partagerai pas, je ne les utiliserai pas, mais je suis intéressé. Mais oui c'est un des professeurs qui l'utilise depuis très longtemps avec nous. C'est un des leaders prof qu'on a depuis le départ. Désolé, je vais devoir y aller.

Ok pas de souci, merci beaucoup en tout cas.

Bonne fin de journée.

Au revoir, bonne fin de journée.

Au revoir.

3. Interview de Maria Perillo

Hello.

Hello, thank you for accepting this interview. We will start by briefly introducing ourselves. So we are Alice and Pauline, and we are studying management in the Louvain School of Management in Belgium. We are currently co-writing our master thesis on voice assistant. The aim of our thesis is to analyze three use cases where voice assistants can bring added value. So, those three use cases are in-car voice assistance, language learning and remote emergency services. In this interview, we will focus on the use of voice assistance in learning a new language. So maybe we can leave you the floor to introduce yourself briefly, your background and maybe your link and expertise with voice assistants.

Okay, I think the most important part of my career that is linked to these voice assistants is when I worked for the MIT, where we developed the first English level test with Alexa. Think about the fact that we did it four years ago, when the smart speakers like Alexa at this assistance, were be used, mostly or 100% for e-commerce. Thinking about moving to education was an innovation, four years ago. What we did is an experiment in computational linguistics.. I co-founded this startup: a language learning platform for learning English. We had 30 million users. The MIT was invested in using our knowledge and data. That's why we entered and we ended up doing this project together. So, at the MIT, the auto ID lab, which is the lab that invented the terms “internet of things”, for you to understand where we were working. So at the MIT, they are trying to create a standard with voice assistance. They think, and they envision, that all brands and platforms will have a voice in the future. So they are creating this database of voices so that in the future, more and more brands can actually create their own tone and style according to their mission and vision. We did this English level test, as an experiment, to demonstrate that a human teacher can create something more than an English level test with no voice assistance. But the experiment was the level of engagement with the teacher. This is the most important thing. We don't think actually that these kinds of experiments will replace language learning, that's for sure. The experiment and all the effort that we are doing in this research is to understand more and more to what extent this combined machine plus human, in this case a teacher, can help you in the future to better learn a language.

So understanding that we are trying to create the best combination of voice assistant and teacher.

The experiment was to understand the level of English with Alexa. Starting from asking basic questions, not only related to grammar or vocabulary, but also related to interest. Why? Because Alexa, as you know, has this huge database that is learning from you as you use it. So, the more you use it, the more it understands you better and can help you learn better. In what sense? It is not going to replace the teacher, but it is going to inform you about something related to your interests. So, if there's interest, and there's motivation, you will learn better. So, that's the experiment. I don't know, if you want to go deeper into some of the questions related to the voice assistant or the educational landscape in this sector. Or what would you like to know?

We have prepared some questions. Actually, you already mentioned some of the topics we wanted to ask you. So, you talked about your experiment, but do you know what are the latest advances in voice assistant used in language learning?

Well, the latest or the ones related to all the things that can be categorized easily. As you know, language learning is something that has several dimensions. If we talk about syntaxes, or vocabulary, or everything that can be categorized, that's where the machine enters and can really help you. If we are talking about other dimension like intonation, for example, or really answering what you are saying, meaning simulating a conversation. That's impossible right now. I mean, it's impossible because we can only simulate, if you have the best simulator, up to 90-95% of what we are saying. Why? Because the machine doesn't understand the meaning of what you're saying. A machine can only understand the vocabulary, the syntaxes, but not the meaning. She knows the answer, but she doesn't know why the answer is correct. For you to understand, a machine, or Alexa, can't really understand where the sentence ends, and where another begins. So, in this landscape of dimension, 90% is a lot. So, we are progressing a lot. And it's true that Alexa and other voice assistant seems like improving, and they will improve in the future. But there's a 5% of that improvement that can never be achieved right now. Because this 5% is the ambiguity of the language, something related to your cartoon for example, or the ambiguity related to metaphors, or that language that only a human can really understand in a conversation. So, more and more progress has been doing in how we create that context. So, the machine can understand the ambiguity. But we are still in the face of researching and testing.

Anyway, the most advanced things have been doing in vocabulary things and, and how to create a sentence up to B level for example, intermediate, but never alone. So, I'm not sure if your question is: can a machine, a voice assistant, replace language learning teaching? No. But it can improve the way we are exposing a user towards more context of language because if I'm using Alexa, while doing other things, I'm exposed to listening comprehension, which is the most important thing for language learning acquisition. So, the first thing that you do when you learn a language is listen like a child listen to their parents when they have to understand. The first thing is to listen. Then they try to imitate the intonation and the pronunciation, and they don't know how to write yet. So, in my opinion, this language assistance of voice assistance can help if correctly used. Because if I started to increase the exposure to listening, I will certainly increase the possibility to pronounce better.

Alexa Echo show is an example of three dimension of multi-device, because it's not only listening, Alexa Echo show has also a display. You have a screen and with a screen, you can also actually see things. And if you add these three things at the same time: listening, speaking and seeing. If Alexa does not understand you, it will say “no, please repeat”. If you put these algorithms in that sense, that can actually not frustrate but motivate, you can certainly improve the exposure and the motivation. So it is better than if you are with a human being maybe, and you're shy. Studies have demonstrated that sometimes when language learning tends not to improve is because you're shy. And these other things that you need to understand in the 360 degree of language learning acquisition.

So, in your opinion, voice assistant could not be used for self-study. But rather as a complementarity to a teacher?

Self-study depends on the level that you want to achieve. Because self-study is always valuable, if compared with nothing. So, it depends on if you are telling a user that you're going to speak as a native speaker only with Alexa, that is not true. But if you tell him, you're going to improve your pronunciation skills, your listening skills, which are the basis for language learning acquisition, of course, they will understand the value of what you're using. Especially when you are at that level of progression where you need to put something more in the ecosystem. So depends on how well is the algorithm performed, because it's not only Alexa, it's the skill that you create for specific purposes, that can be really valuable. Because Alexa works with skills.

The skill that you create, the skill that more developers can create in the future, with specific purposes, can actually improve the language learning system. Because skills that are created specifically for improving that vocabulary, in that context, with this dimension, with these algorithms with the teacher behind these algorithms, who is actually checking and analyzing and improving. Of course, it can be very helpful. This is the experiment that we did at the beginning. To what extent the teacher needs to be there, and what's the role of the teacher for the future, so that more skills with voice assistants can really be valuable? So, it always depends on how to use the technology and not the technology per se.

You mentioned that the voice assistant can really bring help to learning a new language. So, in your opinion, the real added value of voice assistant in learning language, is that the student will learn more?

The hypothesis here, because there's no study today that can really demonstrate, is that the more you're exposed to listening comprehension, the better. So, it is likely to improve. The more you watch movies, the more you improve. That's for sure because we already have studies that demonstrate this. So, Alexa is a voice assistant, and voice listening is something that is for sure important with language learning. So, for sure, it can improve. Because it's more exposure. To what extent? I don't know.

Regarding pronunciation or intonation, do you think it could help?

Well, it's a bit tricky because Alexa itself has its own intonation. And you can have American or US. So, for English language, it's better than nothing. What I'm trying to say is that, for English language, it is one part of the explanation. For another kind of language, like Italian and Spanish, probably pronunciation is not the most important thing. For English, yes, it is. Because we know that pronunciation is one of the most tricky part of language learning acquisition for non-native speakers of English. So, if you only listen, you have something. If you listen and repeat, you have something more. If you listen, repeat, and then you write, you have something even more. It's how you create this path for the user, the user journey that can be improved. So, the way you design the e-learning product, that can actually be improved.

But the pronunciation for example, in the English language test, that we created with the MIT

was actually examined, because it was a test. So, if you don't understand once, and then you don't understand twice, it's one part of the story. Then, you can understand but maybe you are not able to pronounce. So, there we were examining the students. So, we had the table behind, we had that scoring point behind and at the end of the story, at the end of the test, we were able to tell you, "you are advanced level", "your intermediate level" according to how many mistakes you make. That's why we created this algorithm behind. But in general, if you don't have any path, any algorithm, it becomes difficult for beginners' level to start with Alexa if you don't have any knowledge of English because for sure, it doesn't understand you.

So, you think that it can be also a risk that the sometimes people have different accents? For example, in English, we don't have the accent of a native English speaker. So, do you think that it can be a barrier to use the voice assistant?

It can be. If you don't have any level, if you're a beginner, you can't start talking to Alexa. You need to start to talk to a teacher. Then you can improve or you can have a self-study language where you can improve something and then try to understand Alexa. But as I said before, if you create a skill for Alexa with vocabulary, pronunciation, that's a skill for beginner users of English. So, they improve their pronunciation of certain vocabulary. So, it is very well designed the scope of the platform so the user can really benefit from it.

And do you know where we stand in terms of voice assistance adoption? So do we know if teachers and students are willing to use voice assistant in the future of or maybe already now?

They are actually using it. Sometimes we come after. They have a really adopted something and we are trying to adapt. In this case, we know that in the US, three out of four are using Alexa for e-commerce, for specific purposes. And they are actually having this dilemma of "are we are treating them like humans?", and they are not human. In Spain, or Italy or Europe, I have doubts because the current adoption of Alexa in general is not so deep. Then maybe we can talk about the education. But it's still early here to talk about education, where we don't actually see this penetration for e-commerce. We are trying to see more and more podcast-related style things, because we are more used to listen to podcasts. We are that kind of culture and that's where Alexa can help. For example, where you have podcasts, and you can actually

activate podcast, and they are many examples already here in Europe. But something as sophisticated as the English test that we created, I haven't seen yet in Europe.

So it is more adopted in the US than here?

Yes, for the things that I've said before. For vocabulary things, for podcast, and for the things that we are actually doing in other platforms. Think about Alexa, not only as a voice assistant, but as a way to increase the number of devices that we have at our disposal. So maybe you have an iWatch, and you start doing things on iWatch, then you have an iPad, then you have a mobile. Now you also have an Alexa. So, you have more things at your disposal to actually design things that can be useful.

But education is always tricky to enter because e-learning is now becoming powerful because of the Covid situation. Before, with language learning acquisition, we were actually two or three competitors, big competitors. In Europe, Babel, Duolingo and other platforms for English. Now we have lots of more schools with this hybrid approach, and maybe in the future more possibility to use something as different as Alexa can be for educational purposes.

You mentioned some of the applications like Duolingo, and those other applications. What do you think is the best option for learning a new language? Those kinds of applications really designed for language learning or a voice assistant, like a speaker from Amazon or Google?

I believe in a combination of multiple proposals. The optimal combination is self-study plus human touch. So, if you design a user journey, where you have different touch points. Self-study, because you want flexibility, and you also want to study alone, and there are many powerful things that you can do with a mobile. Because you listen and watch at the same time. So, it's very powerful. But then there are some moments where you need to put your language into practice, where you need to activate your language. And you can do this with a human. So, in an interaction where you have more people in a class, so in a synchronous moment you have this activation of language. So, the combination would be poly-synchronous journey, where you have the best of every piece of learning.

And if you could imagine the future of language teaching, what would it be? Do you think it would be used more and more by students and teachers?

Not as a unique tool, it will always be something more that you can add. But you need a base for really learn a language. Alexa is a combination in this ecosystem of proposals that you offer, that can help improve some part of the language learning but not the overall process.

So, the best combination is to combine voice assistant like Alexa with a learning language at school with a teacher?

To me, the best combination is a self-study course with a teacher that guides you. And in this self-study course, you have different devices: you have a mobile, so you have an application. You have a desktop, if you have something that needs to be done in a desktop. The combination is self-study and guided. So, in this self-study approach, Alexa is something more.

From the student's point of view, what might be the barriers to use the voice assistant to learn a new language?

The penetration of the device, of course, that's something you don't have with a mobile, because you use something that they already know how to use it. So, the barrier is the penetration from one side. The second is less study and critical study that can guides you on how to create new skills. And third, that's very personal, because sometimes you have that humanized voice that can help you be less shy as I said before. There are studies in China that have already demonstrated that with a robot, you will have one part of the class that can actually speak more, because they don't feel like they are judging me. They feel like this is a machine, they are not judging me. So, they make more mistakes, and they actually improve because they are alone. With more mistakes, you can actually improve because with mistakes you improve. But then you also have another part of the students that can actually feel not comfortable with this machine voice. But it's still to be demonstrated with other studies. I think the key of the question is how to use Alexa and not just using Alexa.

Do you see any danger in the use of voice assistance? Do you think it could affect the disappearance of some jobs?

No, a teacher will never be replaced by Alexa, for sure. And these are the things that we are telling many teachers that are actually now using technology for the first time. Teacher will never be replaced by a machine. But those teachers who will use Alexa can potentially replace those who can't.

So, are the teachers willing to adopt voice assistants?

Depends on generation, for sure. Here I think that's more to be studied. I know that there are some cartoons in the US that are more willing, and they are spending a lot of time by talking to Alexa. And there are also many parents that are using Alexa with kids so that they are actually improving something. They are spending time with Alexa, and they don't actually listen to Alexa to because when you give a mobile to a kid, of course, you can observe what they are doing. But there is a point of a what they are doing, we don't know.

With Alexa, maybe the parents are more secure, more safe, because they are actually listening to Alexa to at the same time while they are doing other things. So, it's the way Alexa enters the home that can actually change the way we improve language. Because language learning is not only study language, most of the things are related to how we are exposed or to what extent we are exposed to this new language. Because you can improve English if you make the English language enter your life little by little. If I change the language of my device, it's more exposure to language learning because I'm reading a device in English. So, all the things we actually have more people to be more exposed to language learning. So, this is the not direct effect of language learning. But it's more exposure with a culture that is more and more exposed to multiple language learning stimulus.

We are watching movies more and more with subtitles in English and with the help of Netflix, we are improving English skills. Alexa can be used, for example. The limitation that we have now with Alexa is that you can't actually change the language of the device. I know that they are actually trying to improve this because if you set up the device with one language, you can't change. So that's one of the barriers if you want another barrier. But I know that they are trying to improve this and trying to make things smoothly so you can change the store easily. So, they will have more access to other external stores by changing languages as well. But now, it's not

possible.

Is there any other improvement that needs to be done on Alexa like you've just mentioned?

Alexa is improving a lot as more. The more we use it, the more it learns. So, Alexa is improving day by day. The things that we can do for educational purposes are to try to design with Alexa limitations in mind. And the more we know about Alexa limitation, the more we can actually improve this limitation. One thing that they did with Alexa had at the beginning is that they wanted to make it more human. So, they added something like a laugh or different voice when you were doing something strange or whatever. And the effect was a very creepy voice that they had to replace, because it was really scary. Because Alexa was not acting really well. So, try to use Alexa as it is. It can never be a human or so human as you are. So, don't try to imitate the human because we don't want Alexa to be a human. We want Alexa to be as another device that can help.

I think we've covered all the topics that we wanted to ask. I don't know if you have anything to add or some recommendations?

I don't know, the scope of the thesis maybe to help you more.

We are analyzing whether the voice assistant would bring really a real added value in learning language. This is really the main outcome that we wanted to see, and you mentioned that the added value that it brings, is the help.

Yes, more exposures, more listening comprehension, but it depends on how you design the skills. If you design a skill for specific purposes, of course, you can have, as a vocabulary improvement. But the general question: Can, Alexa, replace? No, for sure, no. Alexa can improve if well used for those specific things.

I think we have all answers. Thank you very much for your time.

If you can send me a copy, I would be really interested in seeing the results of the analysis.

Ok, with pleasure. Thank you very much. Have a nice day.

Thank you and good luck.

Cas d'usage n°3 : Services de secours à distance

1. Interview de Axel Legay

Bonjour.

Bonjour, merci beaucoup d'avoir accepté de participer à cet entretien. Donc on va se présenter brièvement, avant de vous poser quelques questions. Je m'appelle Pauline et j'écris mon mémoire avec Alice. On est toutes les 2 étudiantes en dernière année de master en sciences de gestion à la LSM.

Notre mémoire a pour objectif d'étudier 3 cas d'usage dans lesquels l'assistance vocale apporterait une plus-value. Donc ces cas sont l'aide à la conduite, l'apprentissage d'une langue étrangère et les services de secours à distance. On vous a contacté pour avoir votre point de vue sur le ce 3ème cas et voir comment ça peut apporter une aide ou une sécurité aux personnes qui seraient équipées d'une enceinte connectée à domicile.

Donc je vais vous laisser aussi vous présenter, votre parcours et le lien que vous avez avec les assistants vocaux.

Moi, je suis prof à l'UCLouvain mais ça vous le savez puisque vous m'avez trouvé. Je viens de l'université de Liège à la base. J'ai fait une thèse sur la vérification formelle. Ça n'a rien à voir avec les assistants vocaux, c'est des maths, c'est un peu éloigné.

J'ai travaillé beaucoup aux États-Unis, aussi à l'université de Carnegie Mellon et aussi à l'université de Santa Cruz en Californie. J'ai aussi travaillé un peu en France. À l'UCLouvain, je donne des cours à l'intersection entre l'ingénierie des systèmes et la cybersécurité.

Pour les assistants, il se trouve qu'on a pour l'instant un nouveau projet de recherche avec le centre de recherche Sirris, dont le but est d'évaluer la pertinence d'utiliser des assistants vocaux pour les vieilles personnes dans les maisons de repos, en situation d'urgence aussi. Je dois bien dire que c'est par opportunité que ça arrive. Je pense que vous n'étiez pas au courant de tout ça. On vient de commencer le projet et on a eu un TFE aussi là-dessus cette année.

Ah oui, justement on avait aussi contacté Monsieur Rivière. J'imagine que vous le connaissez ? Il nous a redirigé vers 2 étudiants.

Est-ce Stanley et Donatien ?

Oui.

Et vous avez pu faire une interview avec Monsieur Rivière ?

Non, enfin lui, il nous a directement redirigé vers les étudiants.

Ah oui, c'est la vie.

Maintenant, je pense qu'on peut peut-être aller vers des questions plus spécifiques, en rapport avec les secours à distance. Selon vous, quelles sont les avancées actuelles dans les assistants vocaux dédiés aux secours à distance ?

Il y en a eu un chez Google à partir de 2019. L'avancée principale, c'est l'assistant qui remplace la personne pour appeler les services de secours ou alors il n'y a pas que les services de secours, il y a ce qu'on appelle le cercle social. Les personnes âgées aiment bien avoir un cercle social, c'est-à-dire des personnes de confiance à contacter. Par exemple, votre grand-mère pourrait dire, « je contacte ma petite-fille, si je tombe. J'ai pas forcément envie d'appeler les services de secours ». Donc l'avantage c'est que l'assistant, donc l'intelligence artificielle, peut contacter les secours à la place de la personne qui est victime, qui a un besoin, et donc être un peu plus, voire beaucoup, plus clair sur la situation en cours, de façon à ce que les services puissent réagir en conséquence. Donc pour moi, c'est ça l'avancée actuelle. C'est l'arrivée d'une intelligence artificielle, qui s'exprime à la place de la personne en difficulté. Évidemment, la difficulté derrière, c'est de configurer cette intelligence artificielle.

Et est-ce que vous pensez que les personnes âgées ou les personnes qui seraient plus vulnérables pourraient demander de l'aide en contactant ces secours à l'aide d'un assistant vocal ?

Il faut bien comprendre que parfois c'est l'assistant qui va le faire à leur place. Par exemple, la personne tombe, l'assistant appelle les secours parce que la personne est tombée et ne s'est pas relevée. Ou alors c'est la personne qui l'active. J'ai besoin de quelque chose, je pousse sur un bouton. Par exemple, je suis tombé, je pousse un bouton et ça appelle les secours. Donc il y a les deux possibilités et dans les études, on a des personnes âgées qui sont intéressées par ça fortement. Tout comme il y en a qui ont peur évidemment de l'intrusion dans leur vie personnelle, tout comme on en a aussi qui ont peur de ne pas savoir les utiliser et pouvoir les configurer. On a tous les cas de figure. Donc je pense qu'il y a des possibilités mais il y a encore un certain nombre de freins. Aux États Unis, il y a plus d'études qu'en Europe, évidemment c'est toujours pareil. On voit que les personnes âgées là c'est un peu Fifty-Fifty.

Ok. Donc par rapport à ça, selon vous où est-ce qu'on se situe en termes d'adoption des assistants vocaux dans les situations d'urgence ? Est-ce que vous pensez que les personnes en détresse sont vraiment prêtes à faire confiance à un assistant vocal plutôt qu'à une personne ?

Je pense qu'il y a des opportunités. En Belgique, le marché est pour moi à construire. Il y a des opportunités, mais il y a un manque clair de pédagogie, donc d'explication. Il faut passer plus de temps avec les personnes. C'est le problème habituel qu'on a en Belgique : dans tout ce qui est numérique, les gens ont très peur. On est un vieux pays, dans un sens vieux continent, et les gens ont très peur du changement dans le numérique. C'est encore plus amplifié chez les personnes âgées. Je pense qu'il y a une bonne opportunité, mais qu'on ne passera que si on fait l'effort pédagogique. Pour l'instant, on ne le fait pas assez.

Donc au niveau des freins, pour vous, c'est vraiment la barrière technologique pour les personnes âgées ?

Voilà, c'est la barrière technologique, la barrière pédagogique. On entend dans la presse une kyrielle de bêtises sur les données personnelles. Par exemple, la plupart des choses qui sont dites sont fausses. Donc les personnes écoutent le journal tous les soirs et je veux dire que quand on écoute un journal de la RTB, on n'a plus forcément envie d'utiliser son téléphone.

Donc il faut que ça soit fait de façon un peu plus sérieuse par des professionnels, pas des gens qui ont 10 minutes pour faire un reportage, qui ne comprennent rien à ce qu'ils disent. Donc là, on a vraiment un gros problème. Mais je pense qu'il y a vraiment des opportunités et des opportunités pour l'autonomie aussi des personnes. De toute façon, c'est vers cela qu'on se dirige. Les personnes veulent rester chez elles, elles n'ont pas envie d'aller en maison de repos. La crise du COVID a amplifié cela. Ce qu'on a vu dans les maisons de repos, un enfermement de quelques mois pour une personne âgée qui n'a plus que quelques années à vivre, ça ne lui plaît pas forcément. Donc les personnes âgées sont en recherche d'autonomie et les assistants vocaux, c'est une façon de leur en donner et de les rassurer. Encore faut-il leur expliquer.

Vous pensez que la présence d'un assistant vocal pourrait augmenter l'autonomie des personnes âgées ou des personnes en situation de handicap ?

Tout à fait. En situation de handicap, c'est encore plus clair parce qu'un assistant vocal qui est connecté à la maison peut faire beaucoup de choses plus vite que la personne et donc peut lui faciliter la vie. Il y a même l'assistant vocal qui transmet les données médicales de la personne de façon journalière au médecin, qui ne se déplacera pas tous les jours. Il aura tous les jours, le tableau médical de son patient, ce qui est de nature à rassurer le patient. On peut même imaginer des techniques d'intelligence artificielle qui, lorsque le tableau médical n'est pas bon, détecte les anomalies. Donc là, l'assistant n'est plus vocal, mais c'est de la transmission. Par exemple, en cas d'un tableau qui serait un peu bizarre (par exemple, l'augmentation de la fièvre) l'assistant vocal peut interagir et dire "le patient aujourd'hui était dans une telle situation, il a fait plus chaud et il n'a pas bu assez". Donc donner des renseignements aux services de secours pour évaluer la situation. C'est intéressant, donc il peut vraiment se substituer à la personne de façon très concrète et donner un certain nombre d'indications pour décider de la suite.

Donc pour vous, il y a clairement une plus-value au niveau du service de secours mais ça peut aussi être au niveau du personnel médical ? Ça peut aussi aider, par exemple, les aides-soignants, etc. ?

Ça peut aider les aides-soignants et services de secours, ça lie aussi au médical. La personne est tombée, la personne est agenouillée depuis une demi-heure, elle ne se relève pas, elle a poussé un bouton qui dit ça et l'assistant vocal l'explique. OK, on voit en plus que l'assistant

vocal dit que la personne n'a pas pris toutes ses pilules aujourd'hui, ça explique peut-être la chute. Il n'y a peut-être pas besoin d'envoyer le SAMU.

Et est-ce que vous pensez que tout ça est déjà assez implémenté pour que ce soit efficace ?

On a des difficultés de reconnaissance vocale, de configuration donc d'interaction entre la personne et l'assistant. Les 2 étudiants ont fait un travail là-dessus : comment configurer l'assistant ? Par exemple, si je tombe, est-ce que je veux qu'on prévienne mon père ou mon fils par exemple. Donc il faut que la personne âgée puisse répondre à cette question-là. La personne âgée qui prévient son père c'est un peu bizarre mais on pourrait dire que c'est moi dans l'histoire. Donc là on a une technologie de reconnaissance vocale qu'il faut encore améliorer mais les progrès, cette année, sont phénoménaux. Donc on peut vraiment faire pas mal de choses à l'heure actuelle. C'est de l'intelligence artificielle dite faible, donc c'est-à-dire, que ce sont des décisions systématiques. On n'est pas dans Terminator où on apprend. On est dans la prise de décision automatisée, qui est ce que fait l'intelligence artificielle de nos jours. Elle n'est pas en capacité de s'auto-instruire. Donc ce sont ces assistants vocaux pour lesquels cette question se pose. On se posait la question à un moment donné : comment vont-ils s'auto-instruire ? On pourrait très bien imaginer qu'on ait un assistant vocal qui devienne un nazi, qu'il vous donne toujours des idées d'extrême droite quand vous l'interrogez. Donc là, on est dans l'intelligence artificielle un peu moins faible.

Mais pour le médical ici, c'est vraiment de la décision, de l'optimisation et de la réactivité, donc l'intelligence artificielle qui est derrière n'est pas ultra forte, on en maîtrise ces algorithmes.

Au niveau technologique justement, on avait une question, quelles sont les demandes qui restent encore trop complexes à traiter pour un assistant vocal ? Et aussi est-ce que donc dans le cadre des services de secours à distance, ne serait-ce pas la voix qui serait le meilleur moyen de déclencher le l'assistant ?

Alors pour les demandes complexes, en fait, il faut voir comment marche la reconnaissance vocale. Vous allez parler et ça va devenir un ensemble d'objets mathématiques. Si vous déformez un peu la phrase, on va pouvoir avoir un autre objet mathématique à une certaine distance et donc tant qu'on reste à une distance autorisée, l'assistant vocal va être capable de

s'adapter. Maintenant si on change vraiment la formulation, il faut reconnaître qu'une autre formulation veut dire la même chose. C'est parfois compliqué, donc il faut vraiment l'entraîner, mais ça va de mieux en mieux. Donc vraiment ce verrou technologique diminue.

Maintenant sur le déclenchement, donc est-ce qu'on le déclenche par la voix ou pas ? Ça dépend, une personne sourde, muette, quand elle tombe, elle ne peut pas parler, mais son assistant vocal peut parler pour elle. Elle aura besoin de pousser un bouton pour dire à l'assistant vocal ce qu'il doit dire. Il aura besoin de l'avoir configuré avant avec une personne qui va l'aider à le configurer. Ça va un peu dépendre de la personne.

Donc ça serait un système personnalisable ?

Ça doit toujours être personnalisable donc il faut personnaliser les personnes qu'on veut contacter. Et la 2ème chose qu'il faut personnaliser, c'est la façon d'interagir. Les personnes dites faibles parce qu'elles sont plus âgées, certaines n'entendent plus bien, certaines ne sont plus capables de s'exprimer correctement. Si vous écoutez, même dans la politique, Elio Di Rupo, il ne parle plus comme avant. Donc si l'assistant vocal est configuré il y a 30 ans et qu'on lui parle maintenant, la probabilité que les phrases soient toujours bien reconnues est faible. Donc il faut le ré-entraîner et l'adapter à la personne en permanence. C'est une condition nécessaire et c'est indispensable aussi pour rassurer la personne qui va utiliser l'assistant. Si vous donnez un assistant à la personne et que vous lui dites "configurez-le, répondez aux questions" mais que vous n'êtes pas à côté pour l'aider, la probabilité qu'elle l'adopte diminue. Donc il faut investir dans la personne qui va aider à la configuration, qui sera personnalisée. On doit personnaliser et il faut aider à personnaliser. C'est deux choses différentes.

Est-ce que vous pensez que quand l'assistant vocal sera capable de reconnaître les émotions dans la voix, ça sera une plus-value dans l'assistance aux services d'aide à distance ?

Alors ce sera une plus-value, mais ça c'est de l'intelligence artificielle... C'est très compliqué de reconnaître les émotions. Par exemple, je vais arriver près de vous et je vais faire ça **petite claque sur la joue**. Vous êtes une fille donc vous n'avez pas l'habitude ? Moi non plus, je vous rassure. Mais bon peut-être que quand je fais ça à vous, ça vous fera plus mal qu'à moi, qui ai fait des sports de combat et donc vous allez dire « ça me fait très mal et ça me choque ». Moi

on va me le faire, je vais dire « qui a osé faire ça » mais ça ne me fera pas mal. Donc vous voyez, pour la même chose, le même acte, nos émotions vont être extrêmement différentes. En modèle mathématique, parce que l'intelligence artificielle c'est des modèles mathématiques, c'est compliqué.

Aller chez le médecin, vous avez mal au ventre. Moi, dès que j'ai mal quelque part, je suis persuadé que je vais mourir, c'est toujours pareil. Donc, quand le médecin me demande « est-ce que vous avez mal ? » je réponds « j'ai très, très, très, très mal ». La plupart du temps, même les plus petits disent « non, je le sens à peine ». Si le médecin est en face, qui peut vous comprendre et vous estimer, sa plus-value en tant qu'être humain va apporter quelque chose. C'est très peu quantifiable en termes d'intelligence artificielle, donc si on arrivait à quantifier ça pour savoir si une vieille personne exagère ou pas. Par exemple, ma grand-mère appelait l'hôpital toutes les deux minutes en disant qu'elle allait mourir à chaque fois qu'il faisait noir. Elle n'avait strictement rien. Mais comment le savoir sans lui parler, sans lui poser des questions, sans la connaître ? C'est compliqué et on n'est pas du tout au niveau pour faire ça pour l'instant. Les émotions d'un être humain, c'est quelque chose, et peut-être heureusement, que la technologie ne quantifie pas encore de façon très correcte. Il y a beaucoup de biais. Donc on est loin du compte. Si on y arrive un jour, oui, ça sera une plus-value mais je pense que ça va être compliqué. On a vu arriver des choses avec des biais. Il y a déjà des gens qui proposent déjà des choses mais le taux d'apprentissage et le taux de succès dans l'apprentissage, c'est très faible.

Vous avez déjà parlé de quelques risques, mais du coup, qu'est-ce qui serait vraiment risqué dans le fait d'implémenter un assistant vocal dans une situation d'urgence comme celle dont on a parlé ?

Alors, c'est pas les risques d'un assistant vocal en tant que tel mais c'est les personnes plus âgées. Il y a tout ce qui concerne leurs données personnelles. Ils ont donné beaucoup d'informations, ne serait-ce que les personnes qu'ils connaissent, leur façon de réagir. Ces personnes sont très réfractaires à ça parce qu'elles sont persuadées que leurs données sont diffusées partout voire même que les firmes qui possèdent leurs données peuvent leur faire perdre des droits. Il y en a qui sont persuadées que leur assurance augmentera. Et je ne cache pas qu'il y a des projets en place pour le faire, par exemple. Donc il y a ce premier risque là.

Le deuxième risque, c'est d'éliminer l'humain et d'avoir un assistant vocal qui identifie mal la situation. La personne est tombée mais pour l'assistant vocal, ça ne nécessite pas l'envoi des secours. En fait, la personne est tombée, elle s'est ouverte le crâne, elle est en train de se vider, elle va mourir. Il y a des risques de mauvais fonctionnement. Il y a le fait qu'un assistant vocal peut être *hacké* s'ils utilisent les enceintes de type Amazon et autres. Donc ce sont des risques très concrets sur l'utilisation de l'appareil.

Puis, il y a les risques d'adoption : si on investit là-dedans, qui va l'utiliser. Voilà un ensemble de risques.

Justement, vous venez de parler d'un assistant vocal de type Amazon, vous pensez que c'est mieux d'avoir un assistant vocal via des enceintes connectées, comme Google ou Amazon ou justement un autre moyen, une autre application ?

Je vois que les choses comme Amazon, c'est bien pour la domotique. Alors il faut savoir que toutes ces enceintes fonctionnent avec des conditions d'amélioration. C'est-à-dire qu'ils peuvent vous enregistrer pour améliorer la reconnaissance vocale. Donc, on a eu des scandales par le passé où on pensait que l'amélioration était automatique mais en fait c'était des Indiens qui vous espionnaient. Un Indien, si vous le payez 300€, il travaille deux mois. Amazon a eu ce problème d'externaliser tout en Inde. C'était vraiment de l'exploitation d'êtres humains, mais c'était aussi de l'espionnage très clairement. Certaines études américaines montrent que l'assistant se déclenche tout seul une vingtaine de fois par jour. Donc ça veut dire qu'il peut écouter chez vous une vingtaine de fois par jour. Je ne suis pas un grand fan de toutes ces enceintes parce qu'on ne peut pas les contrôler entièrement. Donc je crois que si on est vraiment sur le secours, sur la vie privée, on devrait être sur quelque chose de plus restreint avec quelque chose de plus souverain. Donc on doit savoir comment se fait l'amélioration vocale, où sont stockées les données, qui a accès à quoi. C'est des choses sur lesquelles on travaille en Belgique. Évidemment, ça sera plus cher mais c'est plus sûr. Pour ce genre de situations, qui sont sensibles, il faut des choses sûres. Donc, Amazon et Google le garantissent. Ou Huawei, je pense qu'ils ont aussi leur assistant vocal, Célia, je ne pense pas qu'ils sont totalement sûrs. Maintenant, si c'est pour ouvrir ou fermer les volets, voilà. Ça dépend de ce qu'on va lui demander.

Donc pour vous, quelle serait la meilleure option, que ça soit développé par qui ? Ça serait peut-être aussi un moyen de lever certains freins à l'utilisation ?

Moi je vois bien ça développé par des petites start-ups belges. Il y en a beaucoup qui ont la capacité de le faire, avec des gens de proximité. Je sais bien que la technologie éloigne les êtres humains. Je pense que là il y a besoin d'avoir des êtres humains dans la boucle. Les personnes âgées ne sont pas nées avec un téléphone dans la main. Il n'y avait pour certaines même pas de téléphone quand elles sont nées et encore moins de tv. Moi je verrais bien des petites start-ups belges avec des gens qui vont sur place, avec des *clouds* souverains, les données et les API entièrement maîtrisés par la start-up et pas de dépendance. Ne pas faire des choses qu'on ne maîtrise pas et le dire. Dire comment c'est conçu, la façon dont les données sont gérées, pourquoi est-ce que ça respecte le règlement général pour la protection des données, donc une transparence totale.

Et donc vous, est-ce que vous êtes plus d'avis que justement, l'assistant vocal va renforcer la sécurité des personnes âgées qui en auraient un ?

Je pense que ça va faciliter le dialogue. Mais je crois qu'à l'heure actuelle, l'être humain est nécessaire. Donc si on dit l'assistant vocal sans être humain, je pense que ça va déformer la sécurité. Si c'est un assistant vocal avec des êtres humains derrière, je pense que ça peut optimiser les choses, être plus réactif. Pour l'instant, on a toujours des êtres humains dans les services de secours. Donc on ne voudrait pas qu'un assistant vocal appelle un service de secours appelle lui-même les services de secours, il serait lui-même une intelligence artificielle. Il y en a qui n'ont pas beaucoup de sensibilité. Il faut qu'un moment donné, si l'assistant vocal appelle une personne, cette personne puisse dire, l'assistant vocal ne suffit pas. Il faut qu'on puisse intervenir sur place très vite.

Et est-ce que d'un côté ça pourrait aussi être rassurant pour les proches de ces personnes de savoir qu'elles sont équipées d'un assistant... ces personnes âgées ou des personnes handicapées ?

Je pense que ça peut être rassurant pour eux et plus c'est évalué et mieux ce sera. Par exemple, j'ai une tante qui est tombée, elle avait un de ces fameux colliers là. Elle s'est pétée les 2 bras, elle n'a jamais su aller toucher le collier. Si on est maintenant donc l'évolution ou le simple fait d'avoir la gravité qui chute d'un coup, identifie un problème, on est dans ce genre de techniques, c'est ultra rassurant. Une personne qui doit taguer ses pilules ou simplement ouvrir le pilulier et que le fait qu'elle ne l'ouvre pas, le l'assistant, le détecter et vienne, ça libère sa journée. On ne doit plus tout le temps penser à aller voir sa pilule. Il y a un moment donné, la voix qui lui dit « t'as pas pris tes pilules, va les prendre ». Donc ça rend les choses plus confortables.

Idem pour la domotique, je suis une personne âgée qui préfère dormir dans mon living, mais j'ai des volets dans ma chambre, je dis « vérifie que tous les volets sont fermés ». Oui ils sont fermés, je ne dois pas prendre le Stana pour monter. C'est un effort en moins.

Tout simplement l'assistant vocal qui regarde vos paramètres et vous vous posez moins des questions sur la santé parce qu'en cas de problème avec votre corps, l'assistant va appeler pour vous. Si vous vous déshydratez ou des choses comme ça, on peut imaginer les assistants vocaux qui sont reliés à des puces dans votre corps et qui peuvent mesurer un certain nombre de choses. Par exemple le sucre dans le sang et toutes ces choses, on peut très bien imaginer et qui va enclencher une action pour vous, ça c'est hyper rassurant. Mon père qui a le diabète, qui ne sait plus comment il doit se piquer. L'assistant vocal le prévient puis il contacte les secours. Oui, évidemment, c'est rassurant. On se sent moins seul. C'est important, ça permet de rester chez soi plutôt que d'être dans une maison de repos où l'infirmière passe et fait exactement la même chose.

Surtout que vous disiez justement que les personnes âgées ont de moins en moins envie d'aller d'être placées dans un home ou dans une maison de repos.

Elles ont peur parce qu'elles ont vu que quand elles étaient placées avec le COVID, qu'elles ne disposaient plus d'elles-mêmes. Donc ça leur a fait un sacré choc. J'ai travaillé sur une application qui s'appelle *Coronalert*, combien n'ai-je pas entendu des personnes de 95 ans me dire « écoute, moi je veux bien sortir 3 mois, attraper le COVID et en mourir mais je veux voir mes petits-enfants ». Cette dame était tout à fait mobile. Enfin de quel droit est-ce qu'on la maintenait dans sa chambre ? Ça m'a posé un problème éthique assez important.

Et ça peut être parce que les assistants vocaux, quand vous sortez d'une opération, vous savez qu'on veut raccourcir le temps passé à l'hôpital. Donc vous pouvez rentrer plus vite chez vous, mais même si vous n'êtes pas encore tout à fait en état, l'assistant vocal peut faire un certain nombre de choses et informer le médecin. Donc ça serait de nature à éviter de rester planté des jours dans son lit, à surveiller les métriques pour savoir si tout va bien se passer. Et ça évite aussi de choper la maladie à l'hôpital parce qu'on en chope de plus en plus quand on est à l'hôpital donc ça a cet avantage là aussi.

Donc tout ça, vous pensez que ça va constituer justement le futur ?

Je pense que c'est le futur. Je vois qu'on va vers là. On va de plus en plus vers ce genre de choses. Aux États unis c'est déjà le cas, on ne va même pas parler de l'Asie parce qu'ils sont encore plus en avance puisqu'ils n'ont aucune loi. Donc ils peuvent tester les choses plus vite que nous. Notre risque, c'est vraiment qu'avec des lois trop contraignantes et de la démagogie, on répond à des questions d'idéologie avant de répondre à des questions technologiques. C'est extrêmement dangereux. On le voit dans un sujet comme la 5G, on dit c'est pas bien avant même d'avoir pris les mesures parce qu'il y a des gens qui pensent que c'est pas bien, on est un peu là-dedans pour ce genre de choses avec les données de vie privée. Donc il va falloir qu'on en sorte, qu'on pose des questions et qu'on donne des réponses factuelles et pas des réponses de presse avec des philosophes. Mais on y passera, coûte que coûte, ça arrivera parce que le reste du monde fait ça, même les Africains font ça. On ira vers là.

Et quel serait le moyen pour vous alors de sortir de ce schéma ?

L'éducation. Je pense que dès l'enfance, il faut absolument avoir des cours technologiques. Je vois des enfants de 12-13 ans sur les réseaux sociaux, je vois ce qu'ils mettent, je me dis si on les avait un peu informés, ils ne le feraient pas. Donc dès l'enfance, les informer et puis laisser les générations se lisser, avoir des plus jeunes qui informent les plus vieux. Si les petits enfants sont formés à l'école, ils pourront dire à leurs grands-parents « écoute ce que le journaliste a dit, il a tiré du machin, c'est pas ce mot-là ». Voilà donc c'est intergénérationnel, ça. L'éducation des plus jeunes va aider les plus vieux. Il faut aussi éduquer encore un peu les plus vieux, même s'ils ont déjà du passif. Il faut vraiment attaquer les deux côtés. Je pense que quand on a fait ça

on a gagné. En tout cas, on a réduit ce qu'on appelle la fracture numérique parce que c'est le grand problème de la Wallonie. Il y a en Wallonie un manque d'éducation numérique qui est très fort. Ce qui ne veut pas dire que les gens sont bêtes parce qu'il y en a qui disent que s'ils ne savent pas se servir d'un ordinateur, ça veut dire qu'ils sont cons. C'est pas du tout le cas. Ça veut simplement dire qu'on ne les a pas mis dans cette situation-là. Il faut qu'on change ça, tout en gardant ce qui fait de nous un peu plus intéressant, c'est la partie humaine. Je pense qu'il faut trouver le bon *trade-off*.

Je pense qu'on a fait le tour de toutes nos questions. Peut-être une dernière, comment voyez-vous le futur des services de secours à distance ?

Dur à dire. Je vois les assistants comme des aides qui permettent de répondre à des actions ponctuelles de *monitoring* ou des actions d'urgence en poussant, disons sur un bouton et laisser l'assistant faire le reste avec des êtres humains derrière, qui assistent cet assistant. Les services de secours en Belgique, il faut encore les faire évoluer. Ils sont très peu informatisés et sont très peu soutenus. Toute la numérisation du ministère de la santé en Wallonie a durci, mais c'est colossal quoi.

Et pour vous, ça n'amènera pas une disparition de certains métiers ?

Ça dépend. L'intelligence artificielle faible, c'est vraiment une optimisation de la réactivité. Vous prenez du temps pour réfléchir et donner une réponse. Dans des situations très concrètes, il va le faire plus vite que vous. Donc si c'est très systématique, ça peut remplacer. Comme l'assemblage à la chaîne a remplacé l'être humain, ça peut remplacer certains métiers, sauf qu'ici, on est dans un secteur sensible. Donc je ne suis pas sûr qu'on puisse se l'offrir autant qu'on le voudrait. Voilà, mais je suis sûr qu'il y en a qui vont essayer d'optimiser leur profit. Mais je ne pense pas que ce soit une bonne idée. On parle de quelque chose de systématique, mais qui après, suivant la situation, demande les émotions humaines, donc c'est un peu dur de s'en débarrasser. Sauf si on considère que nos vieux sont du bétail et qu'on les traite à la chaîne comme certains pays le voient comme ça. Il ne faut pas oublier qu'un jour on sera vieux donc...

Pour vous il y a toujours ce besoin de l'être humain derrière ?

Oui, je suis persuadé que la technologie peut créer plus de jobs qu'elle n'en supprime, à condition que l'objectif soit de maximiser le confort des êtres humains et pas le profit. À Silicon Valley, on maximise le profit des actionnaires et on fait des choses qui se pètent la gueule. Alors qu'en fait, il faut trouver un compromis et le compromis maximise nombre de jobs. Mais le coup, c'est que les actionnaires qui ont investi ont touché. C'est la question qu'on doit se poser. Les informaticiens, généralement, sont persuadés qu'il faut quand même garder beaucoup d'humains dans la boucle. Peut-être pas au début de leurs études mais s'ils ont fait quelques années, s'ils ont eux-mêmes subi leur création, ils se rendent compte que c'est bien de mettre l'humain dans la boucle.

Ok, merci beaucoup. Je pense qu'on a fait le tour. Est-ce que vous pensez que ça serait intéressant qu'on contacte les étudiants que vous avez supervisé ?

Je pense que c'est intéressant que vous les contactiez parce qu'ils ont vraiment fait l'expérience avec les personnes âgées. Donc, ils vont pouvoir vous expliquer comment ces personnes l'ont perçu et les difficultés d'apprentissage et le besoin d'avoir un être humain, qui assiste l'assistant et la personne. Sur cet axe-là, c'est intéressant parce qu'ils ont vraiment fait l'expérience. Puis ça leur fera plaisir. Quand on fait un travail, on aime bien pouvoir le valoriser.

Oui c'est sûr. On prendra contact avec eux alors.

Ok, si vous avez d'autres questions, n'hésitez pas.

Super merci beaucoup en tout cas.

Merci bonne journée, au revoir.

Oui, bonne journée, au revoir.

2. Interview de Ahmed Bouzid

Hello.

Hello, thank you for accepting this interview. So, we will first start by briefly introducing ourselves. So, we are Alice and Pauline and we are studying management at the Louvain School of Management in Belgium. And we are currently co-writing our master thesis on voice assistance. So, the aim of our thesis is to analyze three use cases where voice assistance can bring added value. So those three cases are in-car voice assistance, language learning and remote emergency services. So, in this interview, we will focus on the last one, the use of voice assistance in emergency situations at home. So maybe now, we can give you the floor to introduce yourself briefly and your background and maybe your link and expertise with voice assistants.

Okay, thank you for that introduction. So, my name is Ahmed Bouzid, founder of a company called Witlingo, based in Washington DC, US. And what we do is we work with brands to help them launch voice experiences, like Alexa and Google. And, we're doing some stuff in social audio to enable the brands to help to enable their own customers to engage them using voice. So, we launched in 2016. Up to this point, our main work has been in conversational technology. So, people talking back and forth using language, using their voice and using language to talk to assistants, mainly Alexa and Google, to retrieve information and to do things, control stuff, or already skewed actions and so on. That's what we do at a high level.

Ok, thank you. So maybe the first question in relation to remote assistance is: what do you think are the latest advances or devices in voice assistance for remote services?

Can you define that for me? Give me a use case.

Yes, it's how we can help people at home in case of an emergency.

Okay, give me an emergency.

For example, if someone falls, how can the voice assistant see it, detect it and help the person? Maybe by calling the emergency or calling a relative...

Ok. And your question to me was, what do I think about that? What's the latest in that case?

Yes.

Ok. So, we've done, and we are doing work to help all the folks who are lonely. So, the social isolation use case is the one that I'm most familiar with. I know, it's not an emergency, as you know, in terms of there's an event that happened, and it's reacting to it. It's more of a state. So, they are lonely, they are socially isolated. How can we use voice assistants to help them deal with that? So, the voice assistant has two components. One is the fact that it enables the user to use voice to engage with it, as opposed to having to go and grab something like a smartphone and touch it. So that's one component. And the other component is the audio. Meaning the content that they hear is audio, therefore, stuff that can be delivered to audio, like emotion, humor, can be delivered, as opposed to non-audio. But it's not easy to deliver that through text, for example, or even images. So voice is obvious. For example, "hey tell me a joke". So, there are two components that it's easy to ask. And when you hear a joke, it's far more effective than reading a joke, right? Because the sound component is there.

So that's unfamiliar with that context, where we're helping folks feel better. And also, not only feel better in terms of consuming content that is mind broadening, that makes you feel positive, that makes you feel connected because you hear the news. But also, being able to engage with other people, connect with other people. As well as be aware of events nearby, so that you can go out of your shell now that the pandemic is winding down, and go and meet with folks and say "So what's happening here? There is an organic food market every Saturday morning, from 8am to 2pm. Oh, I didn't know that." So now, you go outside and meet people and do things that come back and be much better. So, the discovery of what's happening nearby. Basically, what's happening that you can engage in could be physical, or it could be that you find out that there is a zoom, for folks who are reading a certain book.

So, I don't know if I'm answering your question. But it's not an emergency, I don't know how to define an emergency, or maybe you want to broaden it is a situation where somebody is not

feeling good. And it's over a period. And so that lends itself well for voice assistance, especially for the older folks, because they usually build a relationship with the device a lot more than people who use it functionally. I just use it transactionally. "What's the weather", I just use it for things like that. But the older folks will discover they have much more of a relationship with it. So, they consider to be like members, especially if they're living alone. And it's the only voice that they hear, it's the only voice that hears them on a regular basis, multiple times during the day. So yeah, I would say my answer to you is, I don't know much about emergencies in the narrow sense. But I know quite a bit about helping folks who are in a difficult situation.

Ok thank you. Do you think that the elderly or vulnerable people could ask for help and contact the emergency services more quickly by using voice assistants?

No, I think that's definitely a great use case. Have you ever spoken to the person who launched "Ask my Buddy", like when somebody falls? It's a skill, I think it's "Ask my Buddy", it allows you to invoke it. And then it sends a text message or calls a bunch of people at the same time. Are you familiar with that?

No.

Ok, so you should definitely talk to the person. And I know her very well, and I can connect you with her. She'd be perfect for your specific definition of emergency.

Ok. That would be nice, thank you. Maybe a further question: in your opinion, could voice assistants bring added value to the remote rescue services at home?

Yeah, so I think if it's done well, it's all about the design. I think in principle, it should work. But it's all about thinking about the experience and thinking about what are the different types of situations that you can have? So how about the person who falls too far from the speaker? How about the person who cannot speak loud enough beyond a certain distance as you get older? Or how about a person who lost their voice? I don't know.

But for the majority of folks, I guess I can definitely see the value. But again, the design is very important because you can easily design something that is in principle supposed to do

something, but that doesn't do it well. And especially when you're dealing with an emergency design, it has to be really tight. Because if you build a skill and it's supposed to do certain things, he doesn't do them... I think there's liability that may kick in. So that's something that has to be factored in. If I were to do this, I would do a lot of research in terms of the use case, and I would hire really good designers. I will do a lot of testing and I will do releases very slowly. And then I would learn from the release, and then I'll go. I would say the potential is there. It's tough to do it well, because you're dealing with a very important situation, an emergency. So, the consequences of failure are big when it's not done well.

Ok, so you've just mentioned some of them. So, in your opinion, what are the risks of implementing voice assistance in an emergency situation?

I think some of them are just regular skills and actions. Is somebody going to use language that you have not planned for? So, they might say "I am feeling very anxious" or "I think I am having a heart attack". So just number one is the language; how are people going to express themselves? And then number two is what is an emergency? Is it like when somebody is having a physical ailment or something like "I am lonely, I can't take it anymore"? Is that an emergency? So, there is a spectrum. So, I think, the challenge is defining what is an emergency and also defining the language and people expressing themselves.

And then number two, I think reacting fast enough. So, it's an emergency, what are you going to do about it? Are you going to call 911? And that is dangerous because 911 is a liability. What if you don't call 911? What if you try to call 911 and WiFi goes out? So that is another sort of challenge is to ensure that the transaction happens.

And then I would say this device is communal, meaning it's not like a smartphone where it's yours. So, it's communities in the house. Does it need to know who is saying what? For example, if it's one person who is asked for an emergency, then call this person. And let's say it's not life-threatening, let's say it is an emergency where it means that they need to talk to a specific person, "I need to talk to my psychiatrist", for example. What if there are two or three different people there? Does it know the difference between one or the other? That's probably an edge case. But you may want to think about the fact that it's a convenient device, and how does that come into play?

And do you think that it's already quite implemented? The fact that the voice assistant can recognize some trigger words or recognize an emergency situation even if the person doesn't say "Okay, I need 911" or something?

I suspect that there is not something out there that is trying to cover as many people as possible. There may be specific ones, some homes that have been built just for their residents. Which you can do if people are using Alexa for Business. So, Alexa for Business allows you to build a skill only for a certain number of people, or some people opt in. So, I think there may be things like that, and I'm not aware of them. But I would say it would be very bold for someone to launch an emergency skill that is trying to catch emergencies with all the liable that entails.

I think Amazon cells, I don't know if you can call 911 with their phone, maybe you can. But they're very careful about framing it correctly. This is not for any emergency. I think technically, everything is there. It's all about liability. I think, technically, there's value in trying to help somebody. There are many other devices out there, like something that people put around their neck. If they have an issue, they press the button. When they press the button, there is no ambiguity. It's not like they said something that may or may not be an emergency. They press the button. It's all very tested, and wired knowledge goes to a very specific thing. So, the liability is a lot narrower. But yeah, I hope that answers your question.

And, in your opinion, is the voice assistant more useful for emergency situations, or to avoid loneliness, as you already mentioned?

Meaning, do I think that it's more useful in one versus the other?

Yes.

Well, I think that all I was saying is practically speaking. Not in a perfect world. Practically speaking, it is easier to implement something that helps you be less lonely, than it would help you to take care of an emergency. Taking care of an emergency is clearly more valuable, because if it helps you then it has done something very important. But again, the practical aspects of it come in the way. The fact that people are going to be liable, and so on. So, I think

the low hanging fruit here is helping people with stuff like connecting with other people, and feeling less lonely, and being entertained. Those are the low hanging fruit. But it doesn't mean that it is more valuable than an emergency. It is just that dealing with an emergency is more difficult. I think it's solvable. I don't think it's a technical issue. I think the technical issues that I mentioned are all doable. I think it's the liability aspect of it.

So, it could be that if you want to use this skill for an emergency, you need to sign this waiver. This waiver that says, “we're doing the best that we can, but there are times where it might not work perfectly for whatever reason”. So those kinds of things have to come into the mix for something like that you will be deployable.

And do you think that the presence of voice assistants at home can make people feel safer?

Yes, I think safer than if it wasn't there, yes. Because it's additional. So, the answer I would say is yes, it would make them feel that they can do things that they couldn't do before perhaps. It enables them to help themselves. So maybe they're not capable to turn on the lights on their own or to turn off the lights on their own. So, they feel safer that way. So, if there is something downstairs, maybe watch another light before you go downstairs. So, it is safer in that sense as well.

And do you think it will also increase the autonomy of seniors or people with disabilities?

Sure, definitely. So, you don't have to ask somebody to do something for you all the time, like playing the music. It powers you to do things on your own, using your voice, assuming you can use your voice, but yes definitely.

Because moreover, we have seen that people don't want to go to rest houses anymore. They prefer to stay at home when they are old, rather than going to rest houses.

Yes, so you can live independently for a longer time. As long as you can do certain things on your own. I would say I think that's one of the main use cases for seniors, it enables them to be self-reliant.

And where do you think we stand in terms of adoptions of voice assistants for emergency situations at home? Do you think that the people in distress or with disabilities are willing to trust a voice assistant more than a person?

For?

For emergency situations at home, so like we explained, like if they fall, or if they feel that they have some kind of problem?

I think, obviously, you have the choice between a human and a solution other than the human at this point, but if your choice is between nothing versus voice assistants, I think most of them will choose voice assistants. Or if you can have a person with you all the time, then you can have the person there, when you're doing things that can be risky, like take a shower, or a bath, and so on. But when you don't have the person there, you can be more conservative, and not do those things. This allows the person to maybe schedule their day, so that when there are people around them, they can do those things that are risky. And when they're by themselves, they can be a bit safer in their behavior.

And do you think we should keep some human beings in the loop? And maybe also to avoid some jobs disappearing?

I think the answer to the question of jobs disappearing is a bigger question. Meaning, I know in Europe, they are more interventionists. So, they say “ok we need to make sure that certain jobs are kept safe so far”. But then you introduce the problem of “how about somebody who cannot afford the person, they don't have the money?”. Then this choice for them is between not having a human and nothing? Or should there be a choice that we can have voice assistants? So that's why it makes the question not easy to answer. Because we're helping people keep their jobs. But are we also hurting people who now cannot afford to be by themselves all day long? Because we have intervened, and said, “we're not going to have these voice assistance”. So, in my opinion, the way things work out usually is that there's a transition phase. If it's like all of a sudden, a lot of people lose their jobs, it's not good. But usually, that takes over time. It is over time that these things happen.

But yeah, my answer really is that we have to look at the many folks we get touched by these decisions, and not only by one. Not only the workers but also the users. And also, not only the users, let's say the elderly, but also their children have jobs to do. And it's an imposition on them when they have people who get depressed, because they have to take care of somebody all day long. And so, they are not able to be fulfilled in their life. So, it's complicated. People who are being replaced, people who need help, people who are helping those people, and all that ecosystem that needs to be looked into.

Also, do you think that the technology barrier is a barrier to the use of voice assistance for older people? Because maybe some older people are not used to using the technology and so maybe it can represent a barrier to use voice assistants at home?

Yeah, I think that's for sure. I think that voice assistants are, in theory, a lot easier to use than smartphones, when you're a certain age. The vision, also your tactile as well. So, it becomes really harder and harder to use a smartphone. And then, voice assistants, when it's performing, I think it's great. But there are some barriers that I think should be solved. And those deficiencies that should be solved would benefit not only folks who are older, but everybody. An example would be, have you guys ever used the Echo? Alexa?

Yes.

I'm going to show you an article that enumerates the deficiencies. But one of the biggest barriers to high usability is the fact that it doesn't allow you to pause for more than two seconds when you're talking. So, if I say "Alexa, I want to... I want to listen to a song by... let me think... by The Beatles". You can't do that. You have to prepare yourself to say "Alexa, I want to listen to a song by the Beatles". You have to prepare yourself. Now for older folks, it's a real thing. But every time they start talking, and they pause for a second, Alexa says "I didn't understand". And so that's one barrier.

And I think the second barrier is that it doesn't allow you to speak for more than 10 seconds, for example. So, we start talking and then it's enough. I've had enough. Now I'm going to answer you. But I didn't finish. 10 seconds is all that you have. That is the second barrier. So,

I'll send you the article if there are other things that I mentioned. But there are definitely certain barriers.

And they're not only for the older folks, but for anyone it is a pain to engage with Alexa today because she's so sensitive to pauses. And people don't think of language. And when I'm talking to you, I'm not thinking I'm going to say exactly this. Talking and then pause and then say "mmh" comes naturally. You guys don't think that I stopped talking because I paused for two seconds. Imagine talking to somebody, and every two seconds they swipe. And ask, "can you finish your sentence?". It's annoying.

Okay, so endpointing is the biggest barrier?

Yeah, exactly. So, endpointing for both the end, but sometimes also the beginning. Because you have times where you didn't mean to say "Alexa". It just thought you said Alexa. An example I would give is, once I said, "black stuff". I was talking about over cooking the burgers and there was some black stuff on top of it. So, I was talking to my wife and said, "let's scrape that black stuff". And then Alexa thought I was talking to it. So, it came. I said, "let's get that black stuff of the meat". So, it heard "Alexa of the meat" and it started talking about meat. And so that's end-pointing at the beginning. When do you start listening and then when do you stop listening? So those are two issues.

And do you think that it is going to improve over time? That we will be able to talk to voice assistants like we talk to a person?

I think it's very doable. I think we as consumers need to complain. Like for example Amazon. Like I read articles and say, you saw the far fill and it's been seven years. What have you been doing all these seven years? You're building more skills, building more features. Why don't you solve the UI issue first? And once you have that, then you can go further. Because these people have tens of thousands of employees. What are they doing? This is a very important problem, stopping in the middle of the sentence. That's not nice. I think the answer to you is yeah, it's completely solvable. It could have been solved two years ago, five years ago. And so, will it be solved in the next two years or three years? I think if we make a lot of noise, it will. They'll prioritize it. But if we don't, which all people don't, they don't complain all the time. They

always think it's their fault. They think "I'm not using it correctly", "I got to speak faster". They don't say to themselves "this should be better". Because technology intimidates everybody. Also, folks who think that they are technologically behind, so they say, "oh it's me".

When these kinds of devices will be adopted by the market, do you think that it will be provided by Google or Amazon? Or we should rather use a specific application designed specifically for this purpose?

You are asking if one should use their private skills or use Amazon or Google skills?

Yes. What is the best option between using smart speakers/ voice assistants from big firms like Google or Amazon or, like really applications designed for the specific purpose at home?

Yeah, I think definitely using specific ones. Because usually those go very deep into the use case. So, the chances of the experience being better or higher. Amazon is not about specializing; it's about hitting as many people as possible. Which makes sense, they are doing their business model. But if somebody's company is all about helping people at home, with the tablets, and they build software for it, and so on. Yeah, I think chances are that they are going to deliver a better experience. And I'm always a big champion of specialized solutions, because usually you're able to deliver a lot of value there, because you've done a lot of research and you understand the problem deeply, and you're committed to it for the long run, and so on.

And do you think that there are some requests that are still too complex for a voice assistant to handle? So, for our case, remote emergency situations at home.

I think the complexity really is in guaranteeing the identification and the delivery, the execution. I think by definition emergency means urgent. So different types of emergencies. Maybe you can give me an example, so I can go beyond generalities like this.

So, for example, someone falls, and the voice assistant should evaluate if the person fell, and it's really urgent to call 911? Or if it's a small thing, and it's not really urgent to call the emergencies.

Okay, so here, what you're saying is, the person fell, but they're not expressing themselves and this is where the assistant is being proactive given that it heard about the fall. Is that right?

Yes. So, for those kinds of requests, are the voice assistants already well implemented to handle these kinds of requests? Or is it still too complex?

I don't know. I think you should talk to that person that I'm going to connect you with. She should either have the answer or know somebody who has the answer. All I know is that technology exists. And the Echo can hear things and classify them, it can classify glass breaking, ... I think it can classify if a person fell, and so on. Now, what I mean, the danger there is the false positive. So, nothing bad happened. And then you trigger an emergency, which is costly, meaning what if the police have to come? So that's costly, and after a while it becomes an issue. So I don't know enough to answer your question in a useful way. But I can connect you with people who could, hopefully.

Okay, thank you it would be nice. And in the future, do you think that when the voice assistant will be able to recognize emotions in the voice, will it be another added value in the assistance to remote services?

Yeah, I think so. I think if it can detect that you are not feeling good today as opposed to, let's say, the baseline. So, it's sort of "hey, this is how you normally sound. Today you are sounding down". Or you're angry, or you're happy. Now, the question is, what do you do with that information? How do you react to it? I think being able to detect it, it's not a big problem, I think it's almost there. But then what do you do with it, with that information? Does Amazon use it to sell stuff? So, I think the complexity is in what to do with it, in my opinion.

And based on what you know, what do you see as the future of the remote emergency services supported by voice assistance?

I think there is definitely a future. I think there's a spectrum of emergency. I think we will see these deployed for the least risky spectrum. For example, where somebody voices themselves and says, "I've fallen" or "I think I'm having a heart attack", where the chances of the voice assistant getting it wrong are lower. So, at the same time, I think there has to be the whole

supporting structure for this to happen. Liability has to be part of it. And training, so people know what to do, had to be part of it. So, I don't see people just buying the device and taking it home, and that's that. It has to be part of an ecosystem.

An assistant is living at home, or somebody who has now enrolled in a program. And that program has training, has backup. So maybe use the assistant, but then if it's not responding, there's also a backup that you can use. I see the future happening in pockets. And I see it happening in pockets, or specialized situations. And I see it having powerful specialized situations where it's low hanging fruit at the beginning. And then I see that growing. And I want to go back to what I said initially, which is that there is clearly value here. So, then the elements of the value are here. But how to go to market is going to be the challenge. But we are already seeing it, like companies that we said, are very specialized in this or that. I think that's what we'll see deployments. I think once you have successes, I think then people will sort of be less hesitant to adopt.

Just like a long time ago, probably before you were born, the internet came. And I remember saying "I'll never put my credit card on the web, it doesn't make sense". But then after a couple years, of course, I bought stuff and it's safe. I think, initially, everyone thinks it is very risky, but then when we saw that actually in very few cases, people get robbed and we should just be careful. Then we have gotten used to it. And I think the same thing is going to happen, it's going to start small. And then, as people become more confident, you'll see more adoption.

And so overall, do you think that the voice assistants will really bring value for emergencies at home like they will that will bring security and really good value?

I think yes, for some low hanging fruit emergency or security and all that. I think it's going to be for a lot of people. But then as soon as it becomes a real emergency, I think you'll have pockets where it's much more controlled. So, I see the value there. But I see it happening in specialized pockets, at least in phase one.

Ok anything else? Because I got to run to another meeting.

No problem. I think we asked all the questions we needed to ask.

Ok, if you have any questions beyond these, feel free to email me.

Ok perfect, thank you!

Bye.

Bye.

3. Interview de Stanley Goffinet

Bonjour.

Bonjour, comme on t'a déjà un peu expliqué par écrit le sujet de notre mémoire, est-ce que tu pourrais te présenter en quelques mots, nous expliquer ton parcours et quelle est ton expertise en matière d'assistants vocaux ?

Oui, ok. Donc moi c'est Stanley Goffinet et je suis étudiant et je termine ma formation à l'UCL en science informatique. Les assistants vocaux, je les connais grâce à mon travail de fin d'études, j'ai dû bosser dessus et voilà, j'ai mon expertise de ça.

On peut passer plus aux questions par rapport à notre cas, en rapport avec les services de secours à distance. Selon toi, quelles sont les avancées actuelles dans les assistants vocaux dédiés aux secours à distance ?

En fait, c'était la thèse de notre travail, comme quoi ils sont très peu développés et du coup ça va arriver dans ces années-ci. Du coup la société pour laquelle on devait bosser pour notre travail de fin d'étude, ils nous avaient proposé d'utiliser les assistants vocaux pour voir si c'est mieux que d'utiliser d'autres systèmes existants et on a prouvé et démontré que oui, les assistants vocaux vont arriver de plus en plus dans tout ce qui est domotique pour aider les personnes à domicile. Et c'est dans l'air du temps.

Donc pour toi, ça représente une vraie aide d'avoir un assistant vocal qui assistent les personnes âgées ?

Oui, surtout pour certaines catégories de personnes, c'est très compliqué d'utiliser les technologies et les assistants vocaux permettent de simplifier cette tâche où il y a juste un dialogue entre des personnes âgées ou des personnes limitées mentalement, etc. Du coup, c'est bien pour ces catégories d'utilisateurs.

OK, est-ce que tu penses que les personnes âgées ou les personnes plus vulnérables pourraient demander de l'aide ou alors contacter les secours plus rapidement en s'adressant à un assistant vocal et commencent à adopter ça de plus en plus ?

Oui, c'est ça. Ce qu'on a fait nous, c'est des règles de sécurité pour les personnes âgées par exemple s'il tombe, je veux que mon fils soit prévenu, etc. Donc on a facilité cette création de politique par la personne âgée. Donc, au lieu de taper sur un écran ou sur des touches, il parle directement avec l'assistant vocal qui crée et édite ces politiques. Du coup, c'est clairement plus simplifié pour eux.

Donc d'après votre étude la voix c'est le meilleur moyen dans ce cas ?

Le meilleur je ne sais pas mais en tout cas c'est mieux qu'une interface graphique réglementaire.

Ok, donc tu penses vraiment que l'assistance vocale apporte vraiment une plus-value pour les services de secours à distance à domicile ?

Oui, oui. C'est sûr, 100%.

Du coup, les personnes se sentiraient plus en sécurité et d'une autre manière, ça pourrait aussi rassurer les proches de ces personnes-là de les savoir en possession d'un assistant vocal qui peut les aider ?

Oui, c'est ça parce que en même temps il le crée avec eux. Après il a besoin d'un gestionnaire pour paramétrer tout ça. Enfin, la personne âgée ne sait pas initier d'elle-même tous ces moyens. Mais une fois que c'est fait, ça marche très bien et les personnes de la famille sont clairement enclines à utiliser ce genre de système.

Tu saurais expliquer un peu comment ça fonctionne dans votre projet ? C'est vraiment le fait de personnaliser l'assistant vocal pour que les personnes âgées puissent l'utiliser ?

En fait, le principe c'est que l'assistant vocal fonctionne par rapport aux paramètres donnés. Par exemple, si la personne âgée dit « je veux que mon fils soit prévenu en cas de chute, je veux

que le voisin soit prévenu en cas d'activité suspecte dans la maison », etc. Au lieu que lui-même doive définir toutes ces questions, c'est l'assistant vocal qui lui pose des questions. Nous, on a fait un algorithme derrière pour qu'il pose les questions les plus appropriées selon les besoins de la personne âgée. Du coup, il va poser des questions, par exemple, les personnes avec qui il a le plus de contact, les actions les plus risquées à son domicile. C'est un système qui pose des questions et la personne âgée doit juste répondre par « oui » ou « non », ce qui est très simple pour elle et voilà, c'est comme ça que ça fonctionne.

Donc ce sera un assistant vocal personnalisé pour chaque personne ?

Oui, c'est ça, le gestionnaire au départ initialisera le système domotique avec la personne âgée selon ses préférences et ainsi de suite. Elle pourra changer d'elle-même au quotidien si par exemple, elle est en froid avec ses voisins, retirer ses voisins du tout de ces politiques.

Du coup il n'y a plus ce mot déclencheur ?

Oui, c'est ça. Enfin, par exemple, pour notre test, la personne âgée devait cliquer sur une touche, etc. Mais ça marche clairement avec tout ce qui est Google Home, Alexa, Siri. Par exemple, elle pourrait lancer un mot qui lancera l'assistant vocal et il exécutera les tâches qu'il doit faire.

Tu penses justement que ça sera plus développé avec des enceintes connectées, qui sont de Google ou d'Amazon, ou alors plus des applications qui sont vraiment adaptées au domicile ?

Ça dépend ce qu'ils veulent. Par exemple, pour notre projet, ils voulaient vraiment que ce soit en local pour pas que les données soient un peu partout pour les GAFA. Du coup pour garder vraiment les données on a dû utiliser des raspberries, des petits appareils, vraiment locaux, qui ne partagent pas avec un cloud, etc. Et ça dépend de la sécurité que veulent les personnes. Mais les deux cas sont envisageables.

Et pour toi, quelle serait la meilleure option ? Que ça soit développé, par exemple par une start-up ou que la personne achète un appareil de Google ou d'Amazon ?

Ça dépend quelles sont les convictions de la personne âgée, ou de la personne. En soi, si elle n'a aucune crainte par rapport à la divulgation de ses données, elle peut clairement utiliser n'importe quel système. Mais si elle ne veut pas partager ça avec tout le monde, elle peut utiliser plus des petites technologies, faites par les start-ups.

Est-ce que tu penses que la présence d'un assistant vocal peut augmenter l'autonomie des personnes âgées ou des personnes en difficulté chez elles ?

Oui, c'est clairement ça, il faut donner accès. Enfin, il faut que la personne, d'elle-même, fasse certaines tâches pour rendre son habitat indépendant et qu'elle arrive à faire ces actions toute seule.

Oui parce qu'on a vu que les personnes veulent de moins en moins aller dans les homes, les maisons de repos et elles préfèrent rester chez elles.

C'était clairement une des demandes de notre projet. C'était la clé. Enfin, au niveau problème-solution, le problème c'était que les maisons de retraite seront de moins en moins utilisées parce qu'il y aura plus en plus de personnes âgées. Et du coup, il ne faut pas les faire partir de leur domicile, mais plutôt rendre leur domicile apte pour leur vieillissement.

Selon toi, quels sont les plus gros risques d'implémenter un assistant vocal dans une situation d'urgence comme ça ?

Les risques, c'est par exemple, qu'ils ne comprennent pas. Parce qu'il y a beaucoup de problèmes. Par exemple, c'est du *deep learning* tout ce qui est assistant vocal et des fois, s'il ne comprend pas la commande faite par la personne âgée par exemple « à l'aide », elle devra reposer la question et si la personne est, par exemple, en coma ou en convulsion, c'est très compliqué. Il faudrait des sortes de barrières à deux niveaux. Si l'assistant vocal comprend qu'elle ne répond pas, il peut appeler les secours. Mais oui, c'est quand même risqué si ton assistant vocal bug ou par exemple plus de connexion Internet, des trucs comme ça.

Au niveau de la barrière dans l'utilisation, est ce que la barrière technologique peut être un frein à l'utilisation d'un assistant vocal comme ça ?

Oui, clairement. Soit une barrière parce qu'elles ont une crainte par rapport à cette technologie, soit une barrière parce qu'elles ne savent clairement pas les utiliser. On a posé des questions et on a fait un avis avant de faire le projet et il en résultait que certaines personnes âgées n'ont même pas de Wifi, pas de GSM, elles ont rien du tout. Du coup, pour installer de tels systèmes, c'est très compliqué qu'elles les utilisent.

Dans votre étude, vous avez su prouver que dans le futur, les personnes seraient prêtes à avoir du Wifi et faire le nécessaire pour pouvoir utiliser des assistants vocaux ?

Oui, c'est ça. On a même eu de la satisfaction. On a utilisé notre système sur 5 personnes âgées de 70 à 88 ans et il en ressort que sur le *system usability scale* qu'on a eu un score de 88 sur 100 auprès de personnes âgées. C'est pour voir si notre système est vraiment utilisable. Du coup elles sont plutôt satisfaites de ce système.

Et la solution pour contrer ces barrières, ça pourrait être l'éducation ?

Oui, c'est ça. Nous, on est nés dans l'ère numérique, on est nés avec toutes ces technologies. Pour elles, c'est vraiment compliqué, même allumer une télé peut être une tâche compliquée.

Pour le moment, tu penses qu'on se situe où en termes d'adoption de ces assistants vocaux à domicile, pour les personnes âgées ?

Je pense qu'on est au tout début. Je pense que ça va être un effet boule de neige. Par exemple des personnes âgées qui vont l'utiliser, qui vont dire à leurs copains, « c'est trop bien ce truc-là ». Et puis, de plus en plus, mais je pense qu'on est au tout début parce que je connais très peu de personnes âgées personnellement, qui ont de tels systèmes.

Et tu penses que ces personnes vont être de plus de plus apte à faire confiance à des assistants vocaux plutôt qu'à une personne ?

Oui, clairement. Même, c'est plus simple pour elles d'utiliser ça.

Et tu penses qu'il faut quand même toujours garder un contact un peu humain dans la boucle ?

C'était aussi un des trucs que notre prof disait, que notre système ne doit pas remplacer les anciens systèmes, il doit être complémentaire aux autres. Ça doit être un moyen de plus, un outil de plus dans la mallette quoi. Et le contact avec des personnes est clairement important, même pour les personnes âgées.

Tu penses qu'il y a encore des demandes trop complexes à traiter pour les assistants vocaux ? Par exemple, comme tu disais, si la personne à un moment donné ne répond pas, il devrait savoir détecter ?

Traiter tous les cas de figure, c'est impossible. Il faudrait traiter cas par cas et faire un arbre gigantesque pour vraiment dire à n'importe quelle action possible, l'assistant vocal répond mais je pense que ça va devenir de plus en plus simple avec les méthodes d'apprentissage de ces assistants. Ce sera de plus en plus facile de traiter tout ça au final.

Tu as eu des exemples d'action qui n'était pas faisable, des limitations ?

Ça se peut parce qu'on a dû lire beaucoup de littérature avec plein d'assistants ou de systèmes domotiques comme ça et ils disaient que limité par son environnement, on ne peut pas tout traiter, etc. Mais la grosse limitation d'un assistant vocal, je ne vois pas.

Au début, quand tu disais que la voix n'est peut-être pas le meilleur moyen mais c'est déjà mieux que d'autres systèmes ?

Oui c'est mieux que d'utiliser une tablette, un ordinateur, en tout cas pour le cas de figure de la sécurité des personnes âgées. Mais pour des personnes qui connaissent la technologie, c'est clairement plus simple de créer ça sur une interface graphique.

Au niveau de la sécurité, tu es persuadé que ça peut l'augmenter ?

Oui, persuadé. Même le projet de Bruxelles, va être financé par la région Bruxelles capitale. Donc, je pense qu'il y a clairement des attentes et même des envies de la part des autorités.

Tu penses que ça peut aussi être utilisé pour contrer la solitude des personnes âgées ?

Ça peut, clairement oui. Il peut répondre, faire des blagues avec. Ça peut clairement contrer ça, mais ça sera un complément et ça ne remplacera jamais un contact physique avec les personnes malgré tout.

Est-ce que tu penses que l'assistant vocal sera capable de reconnaître les émotions dans la voix ? S'il est capable de reconnaître qu'elle est en détresse, par exemple.

Oui, je pense. Je ne sais pas si ça existe déjà, sûrement. Mais c'est très compliqué parce qu'il y a certains dialectes où on a l'impression que la personne est énervée. Par exemple, en allemand, ça peut vraiment être un langage très soutenu et qui pourrait faire sensation d'agressivité alors que pas du tout, etc. Il faudrait beaucoup de recherches. Mais il y aura clairement moyen je pense dans un avenir proche.

Tu penses qu'il faudrait adapter du coup les assistants vocaux ? Au niveau des langues, par exemple, il y a des grands-parents qui parlent tout le temps wallon.

Oui, il faudra savoir s'adapter à chaque dialecte. C'était aussi une des questions à la fin de notre présentation orale : « que feriez-vous si c'était un truc à implémenter pour un patois flamand ? ». Et on a répondu qu'il faudrait faire des sortes de petites versions à chaque fois pour les assistants, pour chaque dialecte.

Donc s'il était capable de reconnaître les émotions, ça serait une plus-value ou justement ça serait un peu un côté négatif de la reconnaissance vocale ?

Ah oui, clairement une plus-value parce que ce sont des données en plus. Par exemple, si l'assistant sait qu'il y a de la détresse dans la voix, il configurera cet événement déclencheur plus risqué et ainsi de suite. Par exemple, si la voix est très reposée il se dit « ça va, elle est tombée mais elle a pas mal ». Ça sera clairement un plus.

Si toi tu devais imaginer, comment tu imagines le futur des services de secours à distance à l'aide d'un assistant vocal ?

Le futur ? Oui, ce serait un assistant qui saurait capter le sentiment de la personne et aussi refléter ce sentiment. Par exemple, une personne qui ne sait pas répondre, l'assistant pourrait être plus insistant. L'assistant, lui-même, pourrait refléter des émotions pour que la personne réponde plus facilement. Ça serait vraiment une sorte de dialogue avec une personne physique ou il n'y aurait aucune attente et ce serait vraiment fluide. On pourrait l'imaginer sur n'importe quel support, cet assistant pourrait arriver dans le téléphone ou dans les oreillettes de certaines personnes âgées ou des trucs comme ça. Donc sur plus de support.

Dans le travail que vous avez fait, vous avez trouvé une solution au cas où la personne tombe, elle ne sait plus parler ou est dans le coma ?

Oui, en fait ça dépend de la configuration du système domotique avec lequel l'assistant communique. Ça dépend de tous les capteurs en présence. Par exemple, la personne tombe et ne répond pas, s'il y a un capteur dans la maison, par exemple, une caméra de surveillance ou un capteur de mouvement est présent, il pourrait répondre à ce genre de cas. Mais ça dépend de la complexité du système domotique.

Donc pour toi ce n'est pas utile d'avoir seulement une enceinte connectée avec un assistant vocal ? Ça serait plutôt équipé avec une caméra ou autre ?

Oui, c'est ça. Avec son système domotique, l'assistant pourra surenchérir et parler. Mais ça peut clairement marcher sans. Par exemple, pour un assistant vocal d'accompagnement où il doit juste parler avec l'interlocuteur, il n'y a besoin de rien d'autre.

Par contre pour la sécurité, ça doit être un système plus complexe ?

Oui, plus complexe et avec d'autres systèmes. Il y a des capteurs, des caméras de surveillance. Il y a même un thermostat intelligent, un frigo intelligent. Même une sorte de ceinture pour les chutes. C'est avec une sorte de balancier, donc dès que la ceinture se balance, ça détecte que la personne est tombée. Ou encore des systèmes qui peuvent traquer l'activité quotidienne de la

personne âgée. Par exemple, si ça faisait si ça fait 8 ou 9h qu'elle est dans son lit, ça va alarmer pour dire qu'elle n'a pas bougé de son lit depuis longtemps. Mais oui, plus la complexité du système est grande, plus il y a moyen de capter des éléments et de rendre ça, vraiment sécuritaire pour les personnes âgées.

Dans les personnes que tu as interviewées, ils seraient prêts à investir tout ça pour leur sécurité ? Donc, comme tu disais, un frigo intelligent, un thermostat intelligent, ...

Ça dépend. En fait on a posé des questions sur 15 systèmes domotiques ou appareils domotiques pour voir si ça serait utile pour eux. Souvent, des trucs comme le frigo intelligent pour savoir ce qu'il reste dans le frigo, ils n'étaient pas enclins à utiliser de tels appareils. Mais, par exemple, pour des boutons d'urgence, oui. C'est déjà beaucoup utilisé dans des maisons de retraite. Ça c'est vraiment le truc numéro un ou alors la ceinture. La caméra de sécurité intérieure, ils n'aimaient pas trop parce que ça peut être vraiment la vie privée, plutôt extérieure alors. La caméra thermique, ils aimaient bien aussi.

En fait le principe, ce serait qu'elles n'auraient pas besoin d'acheter ces systèmes mais ce serait une sorte de service au mois qu'elle payerait et le service serait instauré autant de mois qu'elle paierait. Comme pour les maisons de retraite.

Ah oui, un peu comme un abonnement mensuel. Ça c'était un des résultats de votre recherche ?

C'était plutôt dans ce que voulaient ceux qui ont investi. Ils avaient dit qu'ils voudraient plus une sorte de service plutôt qu'un kit qu'on vend. Parce qu'il y a aussi une question de maintenance, il faut que ça soit toujours opérationnel.

Il existe aussi des dispositifs qui peuvent rappeler oralement de prendre les médicaments ?

Oui, c'est exactement ça. J'avais oublié mais c'est une sorte de pilulier intelligent. Après les personnes âgées, elles n'aiment pas qu'on leur donne des ordres donc pour le pilulier à voir. Ils disaient « oui parce que j'oublie, du coup c'est bien » mais quelquefois « je n'aime pas, c'est trop contraignant ».

Dans les freins, est-ce que les personnes âgées craignent pour leurs données personnelles ?

Ça dépend. Mais la plupart du temps, elles s'en foutent un peu. Elles disent « de toute façon, je n'ai rien à cacher, ce n'est pas grave ». Mais bon d'autres disent « je ne veux pas qu'on sache ça de moi, je ne veux pas que mon voisin sache le rythme cardiaque de mon pacemaker ». Ça dépendait des gens.

Et toi, tu penses qu'il y aurait un gros travail à faire sur l'éducation ? C'est quelque chose sur lequel il faut travailler pour que ça fonctionne ?

En fait, ça dépend. Ça peut aller dans les deux sens, soit il y a un gros travail d'éducation à faire par rapport aux personnes âgées, soit rendre le système tellement simple, qu'il n'y aurait même pas besoin d'éducation que ce serait inné de savoir-faire ça. Donc c'est soit, soit l'autre ou les deux à la fois.

Et vous, c'est ce que vous avez tenté de faire avec votre projet, c'est que ça soit le plus simple possible ?

Oui, le plus possible parce que souvent les personnes âgées, ne vont pas commencer à apprendre des trucs. C'est tellement plus simple dans cette approche.

Ok, je pense qu'on a fait le tour. Merci beaucoup.

De rien, avec plaisir.

Résumé :

L'assistant vocal va présenter des opportunités futures importantes dans l'aide à la conduite, l'apprentissage de langues étrangères et les services de secours à distance, ce qui renforcera son adoption. Deux des trois hypothèses de recherche élaborées dans ce mémoire ont pu être acceptées. L'une d'elles, relative à l'aide à la conduite, a dû être rejetée. En effet, l'assistant vocal n'apporte pas encore une réelle plus-value en termes de sécurité et d'expérience aux automobilistes. Pour ce qui est de l'apprentissage d'une langue étrangère, l'usage de l'assistant vocal dans ce domaine apporte une plus-value en tant qu'outil pédagogique de support. Enfin, à l'égard des services de secours à distance, l'assistant vocal adopté renforcera la sécurité des personnes en possession d'une enceinte connectée et rendra plus efficaces les interventions en cas d'urgence à domicile.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique

www.uclouvain.be/lsm