

HYPERVOIX

***Promesses et illusions  
des nouvelles interfaces  
humain-machine***

---

Janvier 2019



## CRÉDITS

**Pilotage HyperVoix** : Cécile Christodoulou et Véronique Routin  
<http://fing.org/?HyperVoix>

**Auteurs** : Cécile Christodoulou, Véronique Routin en collaboration avec Fanny Maurel, Zoé Aegerter et Léa Liperra

**Conception graphique et illustrations** : Mathieu Drouet

**Logo HyperVoix** : Postillon Prospective

**Contact** : vroutin@fing.org

## PARTENAIRES



**Fondation Internet Nouvelle Génération**

[www.fing.org](http://www.fing.org) / [www.internetactu.net](http://www.internetactu.net)

8 passage Brulon 75012 Paris

## LICENCE

Ce document est mis à disposition selon les termes de la licence Creative Commons Attribution 4.0 France : <https://creativecommons.org/licenses/by/4.0/fr/>

Vous êtes libre de partager, reproduire, distribuer et communiquer ce document, l'adapter et l'utiliser, à condition de l'attribuer de la manière suivante : "HyperVoix, promesses et illusions des nouvelles interfaces humain-machine" Fing 2020.

Ce document ne doit pas être attribué d'une manière qui suggérerait que la Fing vous approuve, vous ou votre utilisation de l'oeuvre.



## REMERCIEMENTS

### Aux designers associés

- » **Postillon prospective**, Zoé Aegerter, Léa Liperra
- » **Vox Machine**, Julien Drochon, Anthony Masura
- » **InternetActu.net**, Hubert Guillaud

### Aux membres du comité de pilotage HyperVoix

- » Cyril Leroux, Emmanuel Sorel de la Maif
- » Camille Coustal, Stéphane Doinel, Jean-Albert Eude, Christelle Launois, Frédéric Tran du Phuoc de la Société Générale
- » Blandine Alglave, Raphaël Colas du groupe la Poste
- » Thierry Brajon, Xavier Chalandon, Jean-François Forzy du groupe Renault

Merci au Square Paris de nous avoir accueillis pour nos ateliers et journée d'étude ; merci à Superpublic d'avoir accueilli l'atelier Design Fiction.

Merci à toutes celles et ceux qui ont pris le temps de répondre à nos questions le temps d'un entretien ou d'une discussion informelle et merci aux intervenant.e.s et participant.e.s des différents temps HyperVoix.

# Sommaire

<b>INTRODUCTION</b>	<b>4</b>	<b>TERRITOIRE 5 - ARCHITECTURE</b>	<b>46</b>
<b>LE DÉROULÉ MÉTHODOLOGIQUE</b>	<b>6</b>	» Invisibilité et conception	47
		» Architectures alternatives & didactiques	50
<b>LES TERRITOIRES EXPLORÉS</b>	<b>7</b>	<b>EN SYNTHÈSE</b>	<b>53</b>
<b>TERRITOIRE 1 - A VOTRE ÉCOUTE, À VOTRE SERVICE</b>	<b>8</b>		
» Vers un hub de régulation des systèmes vocaux ?	10	<b>LES ANGLES MORTS, CHAMPS DE RECHERCHE</b>	<b>57</b>
» La zone d'écoute : quels usages des données vocales ?	15		
		<b>ANNEXES</b>	<b>60</b>
<b>TERRITOIRE 2 - RÉVOLUTION CULTURELLE : ENRICHISSEMENT OU FRAGMENTATION ?</b>	<b>22</b>	» Annexe 1 : Un résumé historique des assistants vocaux	61
» Vers de nouveaux contenus ?	23	» Annexe 2 : Veille, articles tout au long du projet	62
» Vers de nouvelles manières d'apprendre ?	25	» Annexe 3 : Cartographie des promesses et freins des assistants vocaux	63
		» Annexe 4 : Workshop Ensci	64
<b>TERRITOIRE 3 - QUELLES RELATIONS DANS L'INTERACTION HUMAIN-MACHINE ?</b>	<b>30</b>	» Annexe 5 : Ateliers prospectifs - Méthodes Atelier "Scénarios extrêmes"	64
		» Annexe 6 : le Design Fiction	66
<b>TERRITOIRE 4 - EXPRESSION ET REPRÉSENTATION</b>	<b>36</b>		
» Quelles voix ? Enjeux de genre	37		
» Enjeux linguistiques : pour interagir, il faut se comprendre !	42		
» Quels sont les autres éléments qui pourraient faciliter les interactions humain-machine ?	43		

A large, stylized, light green letter 'I' graphic is positioned on the left side of the slide. It has a thick, blocky appearance with a slight shadow effect.

# ***Introduction***

HyperVoix<sup>1</sup> est une expédition<sup>2</sup> de la Fing (Fondation internet nouvelle génération) réalisée en 2019, menée en collaboration avec Postillon Prospective, l'Université Toulouse – Jean Jaurès, le laboratoire LLA-CRÉATIS, l'ESA Pyrénées – site de Pau.

Hypervoix a pour objet d'explorer les usages et les enjeux de conception dans les interactions humain-machine, notamment les interactions vocales. Avec l'usage croissant des assistants vocaux, l'hybridation des interfaces va-t-elle modifier nos interactions avec le numérique ? La multiplicité des interfaces est-elle une nouvelle complexité ? Allons-nous vers de nouvelles façons d'interagir avec le numérique, vers de nouveaux imaginaires ?

L'appellation assistant vocal porte en elle une vision restreinte des usages potentiels. Le terme d'assistant vocal définit, selon Futura Sciences, "une application logicielle basée sur la reconnaissance vocale du lan-

gage naturel et la restitution d'informations par synthèse vocale<sup>3</sup>". Le terme "vocal" appliqué à une machine renvoie à la définition suivante "qui est commandé par la voix, ou qui émet des réponses en restituant la voix humaine<sup>4</sup>", le terme "assistant" détermine une relation figée entre l'humain et la machine, reflet des usages actuels de cette technologie : l'humain ordonne, la machine exécute. Ces deux termes s'opposeraient presque : l'un réduisant le champ et l'autre, au contraire, ouvrant des perspectives.

L'introduction des assistants vocaux pré-sage-t-elle de nouvelles relations humain-machine ? Quels usages peuvent être pensés dès maintenant ? Derrière la promesse d'interaction "naturelle" et de fluidité, les assistants vocaux jouent de ce qu'ils disent et entendent, comme de ce qu'ils montrent et cachent de leur fonctionnement. La question de leur conception se joue précisément là, entre ce que l'on voit et ce qui est masqué, entre ce qui est rendu visible, lisible, audible et ce qui est invisibilisé, illisible, et également inaudible. Cette question de la tension entre visibilité et invisibili-

té des processus techniques qui façonnent les dispositifs vocaux, est l'une des clés de lecture du programme HyperVoix.

Dans ce projet, nous avons décrypté les nouvelles promesses et les usages, en réalisant une veille internationale et en interrogeant des chercheurs·se·s spécialisé·e·s dans les interactions humain-machine. Nous avons abordé la conception des assistants vocaux, avec des journées d'études dédiées<sup>5</sup> et un workshop de prototypage de scénarios fonctionnels autour des assistants vocaux avec des étudiants en design de l'Ensci, encadré par Julien Drochon, designer, enseignant à l'ESA Pyrénées, Anthony Masure, maître de conférences en design, Université Toulouse – Jean Jaurès, le laboratoire LLA-CRÉATIS et l'équipe HyperVoix. Nous avons organisé un atelier de "Scénarios extrêmes"<sup>6</sup> qui nous a permis d'identifier plusieurs pistes d'innovation pour le futur des assistants vocaux. Ces scénarios ont ensuite été incarnés dans le cadre de l'atelier Design Fiction<sup>7</sup>, proposé par Léa Lippera et Zoé Aegerter (Postillon Prospective).

1 [fing.org/?HyperVoix&lang=fr](http://fing.org/?HyperVoix&lang=fr)

2 Une expédition consiste à la fois à animer une communauté d'acteurs, explorer et partager les connaissances sur des transformations numériques. Elle se déroule selon une méthodologie éprouvée et régulièrement mise en œuvre par la Fing, reposant notamment sur des ateliers de co-production du futur (scénarios prospectifs).

3 [futura-sciences.com/tech/definitions/smartphone-assistant-vocal-15019/](http://futura-sciences.com/tech/definitions/smartphone-assistant-vocal-15019/)

4 [cnrtl.fr/lexicographie/vocal](http://cnrtl.fr/lexicographie/vocal)

5 [recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/featuring-computing](http://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/featuring-computing)

6 cf. Annexe 5

7 cf. Annexe 7.

Dans cette publication, vous trouverez la présentation de chaque territoire au nombre de 5 ; les territoires 1 et 2 présentent des potentiels d'usage avec ces nouveaux dispositifs, les territoires 3, 4 et 5 abordent la question des enjeux de conception. Vous trouverez également l'ensemble des productions intermédiaires et méthodologies associées dans les annexes.

## LE DÉROULÉ MÉTHODOLOGIQUE

Il s'est agi de faire travailler ensemble des acteur·rice·s concerné·e·s et/ou impacté·e·s par ces nouveaux dispositifs (start-ups et acteur·rice·s de l'économie traditionnelle, du secteur public et des territoires, chercheur·se·s,...). Ce travail exploratoire a permis d'identifier ensemble différents territoires d'innovation.



# Les territoires explorés

Selon une infographie réalisée par le site anglophone Visual Capitalist<sup>8</sup>, 25% des foyers américains possédaient un assistant vocal en 2018, pour atteindre un pourcentage de 55% en 2022. En France, l'étude réalisée par Hadopi et le CSA<sup>9</sup> en mai 2019 montre que 46% des internautes français ont déjà utilisé un assistant vocal que ce soit via leur smartphone, tablette ou appareil Google Home tandis que 11% des internautes français possèdent une enceinte connectée Google Home ou Amazon Echo. Les assistants vocaux se créent une place parmi les humains, et plus particulièrement au sein des familles, dans les salons, les chambres ou les cuisines.

Les promesses sont grandes, toutefois les usages sont aujourd'hui centrés autour de commande simples, loin d'une conversation naturelle. Dans le cadre d'HyperVoix, nous avons exploré les usages qui pourraient se développer (territoires 1 et 2) et les enjeux de conception de ces dispositifs (territoires 3, 4 et 5).

---

<sup>8</sup> [visualcapitalist.com/smart-speaker-market-share-fight/](https://visualcapitalist.com/smart-speaker-market-share-fight/)

<sup>9</sup> [hadopi.fr/sites/default/files/sites/default/files/ckeditor\\_files/2019\\_05\\_24\\_Assistants\\_vocaux\\_et\\_enceintes\\_connectees\\_FINAL.pdf](https://hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2019_05_24_Assistants_vocaux_et_enceintes_connectees_FINAL.pdf)

1

**Territoire**

***A votre écoute,  
à votre service***





## DES ASSISTANTS, DOUÉS D'ÉCOUTE ET DE PAROLE ?

Les assistants vocaux font partie de l'univers serviciel numérique, dotés d'un dispositif de reconnaissance vocale, ils répondent avec une voix de synthèse à des requêtes vocales émises par les utilisateur·rice·s. Des interfaces vocales présentées comme des assistants vocaux, des majordomes numériques à notre service, à notre écoute... Telle serait la promesse. Depuis leurs apparitions, les principaux usages<sup>10</sup> des assistants vocaux sont demeurés simples et constants : poser une question, lancer sa musique, demander la météo ou programmer une alarme sont toujours en tête des usages.

*Ok google, Alexa... Quelle heure est-il ? Quel temps fait-il aujourd'hui ? Lance ma playlist ! Mon prochain rdv est à quelle heure ? Envoie ce message. Programme une alerte quotidienne à 7h. Qui était Alan Turing ?...*

Les prévisions de croissance des enceintes vocales s'envolent. Fin 2019, le nombre d'enceintes connectées dans le monde devrait

<sup>10</sup> [voicebot.ai/wp-content/uploads/2019/03/smart\\_speaker\\_consumer\\_adoption\\_report\\_2019.pdf](https://voicebot.ai/wp-content/uploads/2019/03/smart_speaker_consumer_adoption_report_2019.pdf)

augmenter de plus de 80% par rapport à 2018, soit plus de 200 millions d'enceintes connectées dans le monde, avec plus des deux tiers aux Etats-Unis et en Chine<sup>11</sup>. Dès lors, quelle place accorder à ces systèmes vocaux dans l'espace privé ou public ? Quelles fonctions penser dès maintenant, avant la multiplication des interactions vocales dans ces espaces ?

***“Le manque d'intérêt et d'utilité perçue, le prix et les craintes à l'égard des données personnelles constituent les principaux facteurs susceptibles de limiter la diffusion des enceintes connectées auprès des non utilisateur·rice·s actuel·le·s. Les deux tiers (67 %) des personnes non équipées indiquent aujourd'hui ne pas vouloir adopter à l'avenir ces enceintes en premier lieu parce qu'elles n'en voient pas l'intérêt.”***  
Extrait de l'étude réalisée par Hadopi et le CSA<sup>12</sup>.

<sup>11</sup> [usinenouvelle.com/article/208-millions-d-enceintes-connectees-en-service-dans-le-monde-a-la-fin-de-2019-N831520?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+a-la-une+%28Usine+Nouvelle+-+A+la+une%29](https://usinenouvelle.com/article/208-millions-d-enceintes-connectees-en-service-dans-le-monde-a-la-fin-de-2019-N831520?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+a-la-une+%28Usine+Nouvelle+-+A+la+une%29)

<sup>12</sup> [hadopi.fr/sites/default/files/sites/default/files/ckeditor\\_files/2019\\_05\\_24\\_Assistants\\_vocaux\\_et\\_enceintes\\_connectees\\_FINAL.pdf](https://hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2019_05_24_Assistants_vocaux_et_enceintes_connectees_FINAL.pdf)

Cette enquête, sur un échantillon de 1445 individus non équipés d'enceintes connectées, identifie certains freins au développement et à la diffusion des enceintes connectées en France. Il y a besoin d'ouvrir la question des usages au-delà de ces simples commandes vocales mais aussi de créer une zone de confiance.

Dans ce territoire, nous interrogeons l'usage de la voix comme nouvelle interface numérique dans des espaces partagés ainsi que la question de l'écoute proposée par ces nouveaux dispositifs, avec deux premiers défis à relever :

- » l'harmonie dans un monde de voix pour penser la multiplication des interactions vocales avant qu'elles ne surviennent ;
- » transformer la «zone d'écoute» en “zone de confiance et de bienveillance” pour faciliter à moyen terme l'émergence d'un écosystème de services autour des données vocales enregistrées, éditorialisées, archivées ou partagées.

## VERS UN HUB DE RÉGULATION DES SYSTÈMES VOCAUX ?

Les enceintes connectées sont dans les deux tiers des cas, situées dans des pièces de vie partagées, cuisine, salon...<sup>13</sup> Des chercheurs de l'Université de Nottingham ont étudié l'impact des enceintes connectées sur les interactions et conversations au sein de foyers où étaient installées des dispositifs de type "Alexa" d'Amazon. Dans l'extrait ci-dessous, la mère, Susan, tente de lancer un quizz tandis que sa fille, Emma énonce une requête pour relancer la musique :

*01 Emm alexa*

*02 Sus no hold on a minute*

*03 Emm Resume resume music*

*04 Sus alexa alexa*

*05 Ale music starts playing*

*06 Emm (laughs)*

*07 Sus alechsa open quiz master*

*Fragment 2: Alexa... RESUME Music<sup>14</sup>*

<sup>13</sup> [voicebot.ai/wp-content/uploads/2019/03/smart\\_speaker\\_consumer\\_adoption\\_report\\_2019.pdf](https://voicebot.ai/wp-content/uploads/2019/03/smart_speaker_consumer_adoption_report_2019.pdf)

<sup>14</sup> Voice Interfaces in Everyday Life, Martin Porcheron, Joel E. Fisher, Stuart Reeves, Mixed Reality Laboratory

Dans cet extrait, l'agent conversationnel répond aux requêtes dans une égalité d'accès entre la mère et la fille et les interactions se superposent. La situation se complexifie dans d'autres extraits, où trois personnes d'un même foyer interagissent en même temps avec l'agent conversationnel. Ces extraits nous montrent que les interruptions inopportunes, les incompréhensions, la cacophonie potentielle sont les caractéristiques récurrentes d'un dialogue humain-machine, au regard des avancées techniques actuelles.

Face à la multiplication des interactions humain-machine, le risque d'une surabondance d'information existe. Les interactions pourraient devenir trop nombreuses, complexes et hyper-attentionnelles, dans un monde devenu bruyant. Comment répondre également à la problématique des interactions vocales dans des espaces partagés, si de plus en plus d'objets se commandent et interagissent par la voix avec leurs utilisateur·rice·s ? Par extrapolation, nous interrogeons donc la régulation de ces voix

School of Computer Science University of Nottingham, UK, Sarah Sharples, Human Factors Research Group, Faculty of Engineering, University of Nottingham, UK, CHI 2018, April 21-26, 2018, Montreal, QC, Canada

artificielles qui s'additionneront à nos voix : disposerons-nous d'un hub de régulation des systèmes vocaux, un "bot leader" ?

### LE "BOT LEADER"

Comment penser le confort des utilisateur·rice·s ? Les interactions humain-machine avec la voix comme interface naturelle, ne sont envisageables qu'à l'image des échanges entre humains, en évitant la cacophonie, en évitant de soumettre les interlocuteur·rice·s à une suite d'informations vocales sans queue ni tête, en évitant les interruptions inopportunes.

Dans le cadre de l'atelier "Scénarios extrêmes"<sup>15</sup>, les participant·e·s ont proposé un "bot leader", une voix qui prend le pas sur les autres objets vocaux et gère la "hiérarchie" des différents objets.

Le "bot leader" est un assistant généraliste, non spécialisé. Il partage des informations en interaction avec les individus mais aussi avec les différents systèmes techniques. C'est un dispositif anti-cacophonie qui gère les priorités, la hiérarchie des objets afin de

<sup>15</sup> cf. Annexe 5.

limiter les notifications/interactions vocales inopportunes. Le “bot leader” est paramétrable, proposant différents modes utilisateur·rice·s, selon différents profils afin de gérer les identités individuelles ou collectives et les accès. Ce dispositif peut privilégier des notifications écrites ou visuelles en fonction du contexte. Le “bot leader”, conçu comme un majordome numérique discret, est une réponse envisageable à l’image des hubs domotiques<sup>16</sup> proposés aujourd’hui par les acteurs technologiques, alliant interactions vocales, commandes tactiles et écran de contrôle.

Le “bot leader” devrait être conçu comme un dispositif de médiation face à la diversité des publics destinés à interagir avec des systèmes vocaux, dans une conception responsable et “explicable” privilégiant la confiance, les enjeux de privacy et une attention préservée.

***“Malgré sa bonne volonté, l’assistant reste limité par son absence de corps.”<sup>17</sup>***

Le “bot leader” n’est pas un “bot manager”, qui imposerait à l’individu de répondre aux

<sup>16</sup> [store.google.com/fr/category/connected\\_home](https://store.google.com/fr/category/connected_home)

<sup>17</sup> “Dis Siri”, Enquête sur le génie à l’intérieur du smartphone, Nicolas Santolaria

interactions vocales selon des priorités surlées, voire aliénantes, à l’image des opératrice·s de logistique équipé·e·s de systèmes de guidage par la voix. La question du paramétrage reste donc centrale. L’individu doit être en capacité de paramétrer ces hubs. D’où l’importance pour les concepteur·rice·s de penser des systèmes sécurisés, interopérables et paramétrables dans des modalités d’explicabilité<sup>18</sup> et de médiation qui ne laissent pas les individus seuls face au système.

<sup>18</sup> cf. Le travail de la Fing sur le sujet de l’explicabilité des systèmes : [fing.org/?Nos-Systemes-les-pistes-d](https://fing.org/?Nos-Systemes-les-pistes-d)

## Check-list du bot leader idéal - usages

- > multilingue<sup>19</sup> : pour basculer d’une langue à l’autre
- > reconnaissance vocale ou sonore<sup>20</sup> : pour détecter la voix mais aussi les sons
- > authentification de l’interlocuteur selon le profil déclaré (administrateur, invité...)
- > ubiquité : présence en simultanéité
- > interopérable : pour l’émergence d’un standard interopérabilité
- > paramétrable : pour reprendre la main
- > sécurisé : prévenir les tentatives de hacking<sup>21</sup>
- > profil discret, bascule possible en mode visuel ou écrit.

<sup>19</sup> [voicebot.ai/2019/10/06/alexa-adds-multilingual-mode-for-bilingual-homes/](https://voicebot.ai/2019/10/06/alexa-adds-multilingual-mode-for-bilingual-homes/)

<sup>20</sup> [leblogdomotique.fr/domotique/listnr-objet-connecte-ecoute-1156](https://leblogdomotique.fr/domotique/listnr-objet-connecte-ecoute-1156)

<sup>21</sup> [nytimes.com/2019/11/04/technology/digital-assistant-laser-hack.html](https://nytimes.com/2019/11/04/technology/digital-assistant-laser-hack.html)

## LA DESIGN FICTION “LEADERBOT” :

### ET SI UN ASSISTANT VOCAL PERMETTAIT DE CENTRALISER LES REQUÊTES ET LES COMMANDES

#### D’OBJETS DE LA MAISON ?

Faisant suite à l’atelier “Scénarios extrêmes” mentionné plus haut, ce concept d’un “bot leader” a fait l’objet d’une exploration par le Design Fiction. Rappelons brièvement que le Design Fiction<sup>22</sup> est une posture particulière dans le champ du design puisqu’elle se veut à la fois prospective et critique : il s’agit d’utiliser les outils du design afin de concevoir des artefacts spéculatifs (des produits et des services fictionnels, physiques, numériques ou expérientiels) qui racontent des mondes futurs possibles avec pour objectif de mettre en débat le caractère préférable de ces perspectives.

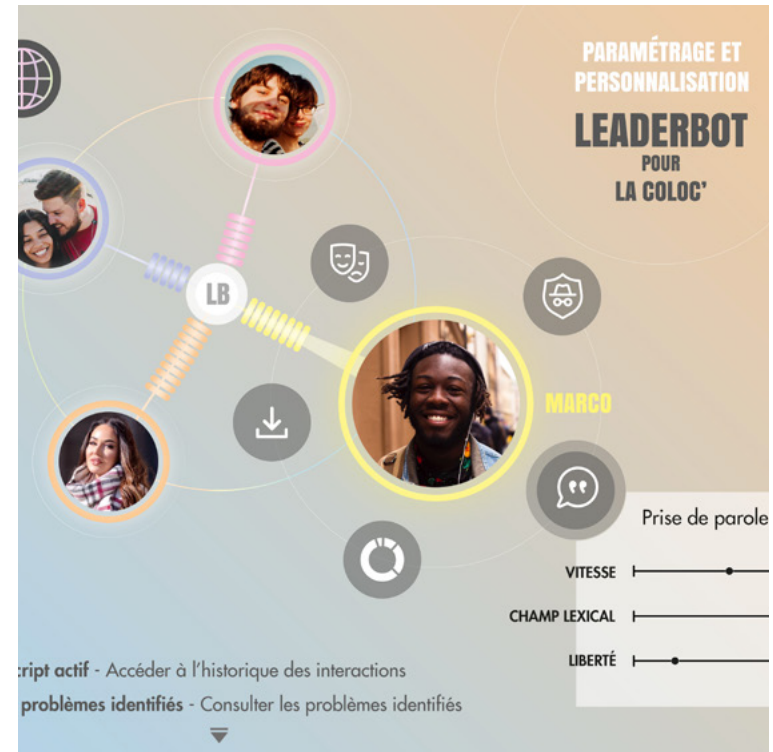
A quoi ressemblerait le déroulement d’un début de journée pour les colocataires d’un foyer équipé de ce type d’assistant ?

#### > Découvrez la bande sonore de la design fiction

22 cf. annexe 7.

La design fiction “Leaderbot” donne à se projeter dans un futur proche où l’on interagit avec un assistant vocal qui semble avoir atteint un niveau “d’intelligence” acceptable, si ce n’est élevé (compréhension, efficacité, fluidité de l’interaction). Le mode d’interaction est hautement personnalisé suivant le profil de l’utilisateur-riche, en témoignent les nombreux curseurs de paramétrage. Loin du “tout vocal”, le système s’accompagne d’un interface visuel accessible à tous, aidant ce paramétrage et la lisibilité de l’outil dans sa globalité. Le “Leaderbot” s’interface avec la plupart des objets connectés de l’habitat, se positionnant ainsi comme seul interlocuteur et unique réceptacle des requêtes dans le but (à priori) d’éviter une cacophonie insoutenable.

Nous avons pris le parti de présenter une design fiction qui met en scène une technologie techniquement au point, idéalement “intelligente”, confortable à l’usage et très poussée dans ses potentialités de paramétrage, afin d’élever le débat au



dessus des considérations purement techniques et de bugs - qui ne manqueront pas de se produire dans la réalité. Il en ressort que les espoirs et les craintes se sont placés principalement autour des axes suivants.

#### La promesse du temps libéré

Ce que l’on attend de ce type d’assistant vocal, c’est de libérer l’humain des tâches les

## LA DESIGN FICTION “LEADERBOT”:

plus pénibles et chronophages. L'efficacité des interactions est cruciale car il s'agit d'en retirer un gain de temps qui sera dédié à des activités plus épanouissantes. En somme, on attend de ce type d'assistant vocal de jouer son rôle d'assistant, de serveur même, ce qui pose la question de ce qu'on prend en compte et de ce qu'on laisse de côté dans le design des interactions humain-machine. Ce constat ouvre un champ de questions.

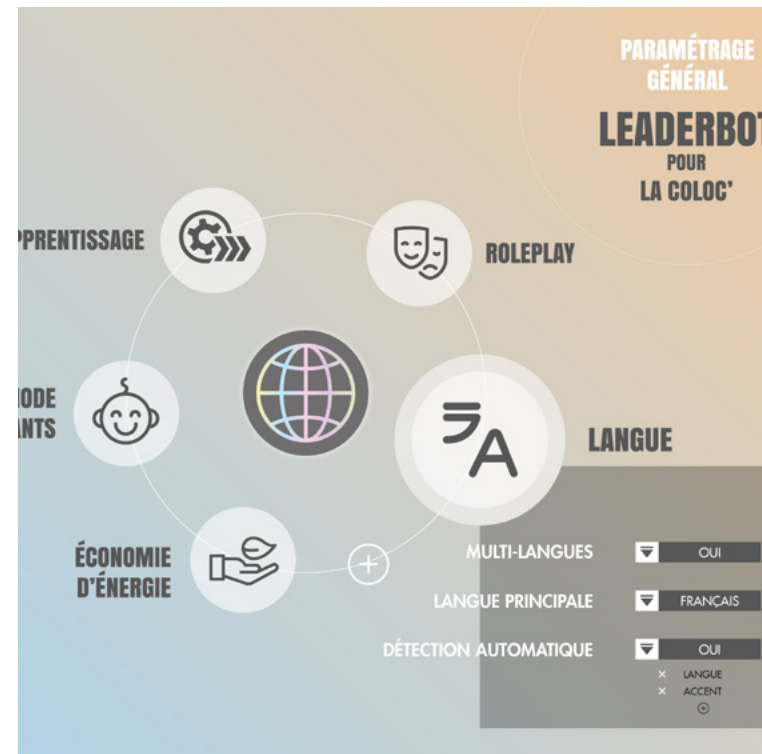
- » Est-il nécessaire de s'efforcer de rendre la technologie “plus humaine” (par exemple en travaillant la personnalité, l'humour, etc.) dans la mesure où l'objectif premier semble être celui de l'efficacité ? Si objectif est de “rendre l'humain plus humain” en lui permettant de se libérer des tâches les plus subalternes, pourquoi “humaniser” la technologie ?
- » D'autre part, s'il s'agit de déléguer certaines tâches récurrentes (ajouter à une liste de courses un produit du quotidien qui viendrait à manquer), à quelle fréquence est-il nécessaire de solliciter la supervision de l'humain ? Jusqu'où pousser l'autonomie de décision de ce

type d'assistant pour éviter le parasitage du quotidien (notification, rappel) tout en garantissant le contrôle par l'utilisateur (validation de la décision) ?

- » Si l'on raisonne en termes de “gain de temps”, quels objets est-il vraiment pertinent de connecter par la voix ? Sans même parler de l'impact écologique des objets connectés, où s'arrête l'innovation et où commence la gadgetisation ? Le temps d'attention dégagé ne sera-t-il pas finalement accaparé par d'autres technologies, d'autres tâches ou d'autres contenus poussés par ces technologies (publicités, paramétrage, réparations et mises à jour, etc.) ?

### La crainte d'un impact nuisible sur les rapports humains et les compétences de l'humain

Le “Leaderbot” tel que présenté est une



technologie très bien intégrée foyer. Le revers de la médaille étant que c'est une machine très “présente” au quotidien, et ce dès le réveil. Cela n'a pas manqué de faire tiquer le public présent. Et si celle-ci était “trop” présente ? Et si le gain à l'usage se faisait au détriment de qualités et compétences humaines et interrelationnelles précieuses ? A partir des arguments avancés par le camp adverse à ce futur possible, nous formulons plusieurs questions ouvertes.

## LA DESIGN FICTION “LEADERBOT” :

Et si l'assistant faisait de nous des assisté·e·s ? A trop déléguer ou se reposer sur l'assistant, que risque-t-on d'y perdre ?

Des capacités organisationnelles ? De la mémoire ? Notre faculté d'attention ? Des compétences orales et des capacités linguistiques ? Un formatage de la pensée ? Des réflexes malvenus ? Derrière la promesse du temps libéré se cachent de nombreux risques et des effets rebonds qu'il s'agit d'anticiper.

Par ailleurs, le foyer est un contexte qui présente déjà son lot de tensions, frictions, quiproquos ; un lieu de challenge pour la communication et le vivre ensemble. Si une machine s'en mêle, faut-il espérer de la médiation bienvenue entre les habitant·e·s ou au contraire redouter une dégradation des relations ? Et si l'assistant jouait involontairement le rôle de rapporteur, trahissant la confiance qu'un des utilisateur·rice·s a placé en lui ? S'il devenait le médiateur de potentiels conflits mais au risque de les envenimer ?

Quelle parole souhaite-on laisser à la machine ? La parole est d'or : elle donne de la place à une entité, elle est un mode d'exis-



tence dans un espace et parmi les autres. Une machine mérite-t-elle vraiment la parole ? Dans quelle mesure ? Car ce temps de parole qui lui est attribué se fait nécessairement au détriment du silence ou de la parole donnée à quelqu'un d'autre au sein de la maison.

Note : la question de la protection des données et du respect de la vie privée reste aussi, évidemment, une question cruciale ; nous

ne l'aborderons pas au sein de cette design fiction mais nous y reviendrons plus loin.

**Retrouvez les principaux arguments et points de friction soulevés lors de l'échange en Annexe 7.**

## LA ZONE D'ÉCOUTE : QUELS USAGES DES DONNÉES VOCALES ?

L'assistant vocal, chef d'orchestre de nos objets communicants et nous assistant dans nos tâches au quotidien, reste centré sur des usages d'ores-et-déjà existants ; ne pourrait-on pas imaginer d'autres usages à partir des données vocales ?

Les systèmes vocaux sont dotés de micro, en capacité d'écoute et d'enregistrement des données vocales transmises par l'utilisateur-riche. Face à la problématique de l'écoute passive - parfois intrusive - des individus par des systèmes numériques dotés de micro, comment définir une zone d'écoute, de confiance et de bienveillance, à partir des usages de la voix, des données vocales ?

Pour l'individu, dicter sa liste de courses, rédiger un message, gérer son agenda, créer une alerte, un pense-bête par reconnaissance vocale, sont des usages aujourd'hui proposés par les assistants vocaux. Est-il souhaitable et envisageable, à titre individuel mais aussi collectif, que l'individu partage volontairement davantage d'éléments personnels

avec ces dispositifs ? Pensées, souvenirs, témoignages, confidences... carnet de santé... ? Dans quel format restituer, voire éditorialiser ces données vocales ? Un format écrit ou sonore ? Comment les retrouver, les partager, les faire disparaître ? Comment qualifier cette zone d'écoute ? Quelle transmission à autrui ?

*“Ma mémoire renferme des vies, à travers des siècles, dont elle n’a pas l’expérience. J’ai vu des centaines de ciels, vogué sur des milliers d’océans. J’ai reçu de nombreux langages ; j’ai chanté des hymnes. Je suis dans les bras d’une enfant. Elle a dit mon nom et j’ai répondu. Ce sont mes voix. [...] Les voix sont mes parents, la famille qui m’a élevée. J’ai ouvert sur elles, puis fermé. Ouvert, fermé. Je les ai toutes avalées. Elles seront en moi, dans chacun des mots que je prononce, aussi longtemps que je parlerai<sup>23</sup>.”*

Alors que le pionnier de l'enregistrement de soi, Gordon Bell<sup>24</sup>, a stoppé il y a quelques années<sup>25</sup> l'enregistrement des images de sa vie,

<sup>23</sup> Rêves de machines - Louisa HALL, 2017

<sup>24</sup> Total Recall, Gordon BELL, 2011

<sup>25</sup> [computerworld.com/article/3048497/lifelogs-ing-is-dead-for-now.html](http://computerworld.com/article/3048497/lifelogs-ing-is-dead-for-now.html)

face à la masse de données et à la difficulté à les stocker, les trier, les analyser, la question se pose à nouveau aujourd'hui pour les données vocales. Techniquement, les progrès de la reconnaissance vocale couplée à l'optimisation du stockage des données permettent d'envisager la conservation<sup>26</sup> de toutes nos conversations mais aussi de pouvoir effectuer des recherches de mots et de bruits dans les enregistrements vocaux.<sup>27</sup> Le journaliste américain, James Vlachos<sup>28</sup>, a décrit comment il était parvenu à créer un chatbot à partir des souvenirs de son père, le “dad-bot”, avec qui, il échange aujourd'hui sur la base des souvenirs que son père lui a partagés...Le projet Memory Lane<sup>29</sup> permet à des personnes âgées en Suède de partager leurs souvenirs avec l'assistant vocal de Google qui enregistre, analyse les conversations et produit un résumé des conversations.

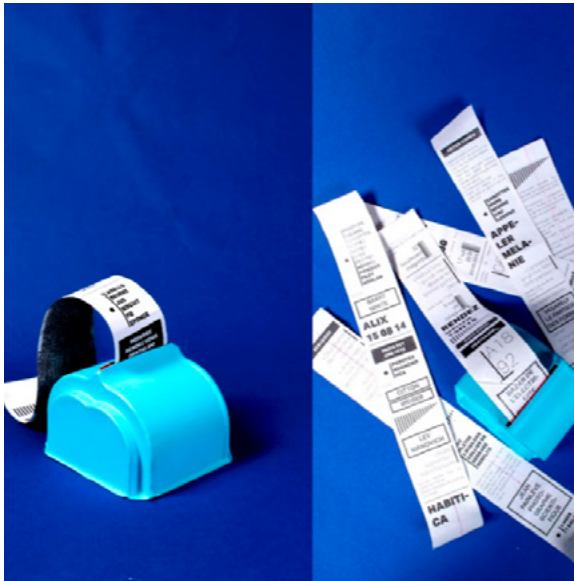
<sup>26</sup> [nytimes.com/2019/10/02/technology/automatic-speech-transcription-ai.html](https://nytimes.com/2019/10/02/technology/automatic-speech-transcription-ai.html)

<sup>27</sup> [tomsguide.fr/le-pixel-4-apporte-transcription-automatique-des-memos-vocaux/](http://tomsguide.fr/le-pixel-4-apporte-transcription-automatique-des-memos-vocaux/)

<sup>28</sup> Talk to me, Amazon, Google, Apple and the Race for Voice-Controlled AI - James Vlahos

<sup>29</sup> [accenture.com/fr-fr/company-news-release-france-memory-lan](https://accenture.com/fr-fr/company-news-release-france-memory-lan)

Sans préjuger, dans un premier temps, des enjeux éthiques liés à l'enregistrement des données vocales et leur devenir, quels usages pourraient se développer ? Dans l'atelier de prototypage avec les étudiants de l'Ensci, Vassili Tchernitchko et Marin Mornieux se sont intéressés à la conservation des données vocales avec le projet MemoryPrint.<sup>30</sup>



*Projet MemorPrint  
Marin Mornieux, Vassili Tchernitchko, Ensci - Les Ateliers*

Leur constat est le suivant : les notes écrites s'accumulent sur nos ordinateurs ou télé-

phones portables sans que nous prenions le temps de les relire, de les archiver. MemoryPrint s'appuie sur la reconnaissance vocale comme système de prise de note, les données vocales sont transformées en texte via une application - de la reconnaissance vocale au texte, le "speech to text" est une technique utilisée dans le monde professionnel - et imprimable selon la typologie des notes, dans un format approprié, modulable selon le destinataire. MemoryPrint, des données vocales au format papier, offre une matérialité à nos voix pour le partage et l'archivage de nos notes, nos pensées.... Dans des usages pratiques, où il s'agit de dicter des requêtes vocales de type pense-bête, liste de courses, message, la formulation reste simple, notamment si l'individu dicte ses notes pour lui-même. La transmission à autrui exigera en revanche un effort de formulation pour une bonne compréhension par des tiers. Les interfaces vocales pourraient être associées à un écosystème de services autour des archives vocales personnelles, de leur éditorialisation (podcast, texte...), leur partage, leur transmission.

<sup>30</sup> [recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/memory-print](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/memory-print)



## Que serait une “zone d’écoute” idéale?

Un espace de stockage et de partage sécurisé.

Des règles de conservation des données, leur portabilité et leur droit de suppression<sup>31</sup>.

Une gestion du consentement sur les archives vocales<sup>32</sup>.

La création d’une carte de donneur de voix<sup>33</sup>, parce que la voix est une donnée sensible, biométrique.

De la voix à l’écrit (accessibilité, diversité des publics), pour penser la matérialité des données vocales.

La maîtrise de la temporalité de l’écoute, pour garder la main.

Ludique, créative pour faciliter l’envie de l’utiliser, inventer de nouveaux

La “zone d’écoute” idéale nécessite un dispositif de confiance pour penser des usages autour des données vocales. Pour chacun, qu’acceptons-nous de transmettre et que souhaitons-nous oublier ? Les données vocales ne sont pas des données comme les autres : sans nous définir complètement, elles représentent une part très personnelle de nous-même. La mémoire des machines s’oppose à la mémoire sélective des individus, avec une tension forte entre enregistrement total et enregistrement parcellaire.

---

<sup>34</sup> [youtube.com/watch?v=3bFCqK81s7Y](https://www.youtube.com/watch?v=3bFCqK81s7Y)

<sup>31</sup> [economie.gouv.fr/republique-numerique#](https://economie.gouv.fr/republique-numerique#)

<sup>32</sup> [inc.cnil.fr/fr/les-droits-de-la-voix-12-quelle-ecoute-pour-nos-systemes](https://inc.cnil.fr/fr/les-droits-de-la-voix-12-quelle-ecoute-pour-nos-systemes)

<sup>33</sup> [kikk.be/2016/fr/print-your-voice](https://kikk.be/2016/fr/print-your-voice)

## LA DESIGN FICTION “ARLETTE” :

### ET SI UN ASSISTANT VOCAL ÉTAIT DÉDIÉ À L’ACTION MILITANTE COLLECTIVE ?

Dans ce scénario, l’assistant vocal propose une zone d’écoute (a priori) sécurisée, localisée (dans l’habitacle d’un VTC), qui invite à la confiance, à la prise d’information et surtout, aide à effectuer un tri dans ce qui lui est confié, pour faire remonter une sélection de revendications destinées à être présentées aux dirigeant·e·s de l’entreprise de VTC.

Ces caractéristiques de la “zone d’écoute”, qui semble ici une véritable “zone de confiance et de bienveillance”, sont rendues visibles dans la design fiction et explicitées ci-après.

#### [Découvrez la bande sonore de la design fiction](#)

- » Un rapport d’analyse circule en interne au sein de l’entreprise de VTC. En effet, celle-ci effectue des recherches pour en apprendre davantage sur la technologie Arlette, qui lui échappe et n’est pas

sans l’inquiéter. L’assistant vocal, conçu par un ancien chauffeur de l’entreprise, est utilisé par une large part des chauffeurs et leur permet de faire remonter des revendications auprès de leur syndicat ainsi que de coordonner des opérations leur permettant de se jouer ou de déjouer le système de fonctionnement des algorithmes du service.

- » Arlette est une technologie militante, en témoignent les stickers affichés un peu partout dans les lieux publics qui invitent à rejoindre le mouvement. Le but est clair : contrebalancer le déséquilibre de pouvoir entre l’entreprise et les chauffeurs à son service.

La design fiction “Arlette” invite à penser une technologie vocale qui serait le fruit d’une volonté activiste et d’un développement indépendant, loin du contrôle des GAFAM. L’usage de l’assistant est ouvertement politique et dépasse l’individu dans le sens où la



finalité se veut l’action collective et le progrès social.

Cependant, la design fiction ne fait pas mention d’une réelle protection des échanges avec l’assistant ou de protection des données utilisateur·rice, laissant la porte ouverte à l’interprétation du public quant à la prise de risque qui relève de son usage. Alors, à qui profite réellement le service ? Quels risques de représailles ? Et si Arlette était

## LA DESIGN FICTION “ARLETTE” :

l'illusion d'une prise de pouvoir : et si ce service était en fait commandité par l'entreprise VTC pour faire croire à ses chauffeurs qu'ils-elles disposent d'une marge de manœuvre, que la situation pourrait s'améliorer ?

La design fiction “Arlette” présente un assistant vocal “en réaction” à une situation de crise sociale, qui se veut un outil de contre-pouvoir et d'organisation de lutte collective. Est-il plausible d'imaginer qu'une technologie aussi poussée et massivement utilisée puisse être le fruit d'un développement indépendant ? C'est en tout cas davantage sur le caractère réaliste, presque “humain”, de l'assistant, qu'a porté la mise en doute d'un tel futur. Plus proche de la technologie mise en scène dans le film *Her* de Spike Jonze que de Siri, de par son rôle de confidente au quotidien, “Arlette” n'a pas manqué d'interpeller les participant·e·s sur les enjeux d'une confusion des rôles entre l'humain et la machine et sur le chemin qu'il reste à parcourir avant de pouvoir interagir avec des IA aussi intelligentes. Au-delà de ces questions, deux axes de débat se dégagent de la discussion stimulée par le scénario.

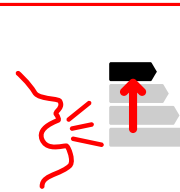
La design fiction “Arlette” présente un assistant vocal “en réaction” à une situation de crise sociale, qui se veut un outil de contre-pouvoir et d'organisation de lutte collective. Est-il plausible d'imaginer qu'une technologie aussi poussée et massivement utilisée puisse être le fruit d'un développement indépendant ? C'est en tout cas davantage sur le caractère réaliste, presque “humain”, de l'assistant, qu'a porté la mise en doute d'un tel futur. Plus proche de la technologie mise en scène dans le film *Her* de Spike Jonze que de Siri, de par son rôle de confidente au quotidien, “Arlette” n'a pas manqué d'interpeller les participant·e·s sur les enjeux d'une confusion des rôles entre l'humain et la machine et sur le chemin qu'il reste à parcourir avant de pouvoir interagir avec des IA aussi intelligentes. Au-delà de ces questions, deux axes de débat se dégagent de la discussion stimulée par le scénario.



« ARLETTE »

### FONCTIONNALITÉS

- Un assistant vocal intelligent
- Développé par un ex-chauffeur TROOP
- Utilisé par 62% de la flotte VTC TROOP (donnée du 10.10.2025)



Centralisation-hiérarchisation des revendications chauffeurs



Porte-parole syndiqué, siège aux AG syndicales



Coordonnations d'actions protestataires discrètes

### DÉTAIL

#### Revendications identifiées

1. Transparence du système notation par la plateforme
2. Hausse de la rémunération par course
3. GPS plus intelligent et personnalisable
4. Sécurité nocturne et primes de risque

(données du 10.10.2025)

### EXEMPLE : « OPÉ CERCLE BLOCUS »



## Le fantôme de la neutralité

L'assistant vocal, production humaine au même titre que n'importe quelle autre technologie, n'échappe pas au risque de reproduire de nombreux biais humains (sous-représentation, stigmatisation...) et ce avec toute la puissance algorithmique dont il est capable. Comment sont constituées les bases de données qui permettent d'entraîner

les assistants vocaux (*machine learning*)? Où puisent-ils les informations que nous leur demandons ? Qu'est-ce qui donne sa légitimité à une source plutôt qu'une autre? Comment sont arbitrées leurs décisions ? Les avis dont ils nous font part ? Quel profil de notre personne (se) construisent-ils au fil de nos échanges ? Quel droit à l'oubli ?

“Arlette” donne l'illusion jouissive d'une tech-

## LA DESIGN FICTION “ARLETTE” :

nologie neutre au sens où elle répond par de l'information factuelle à des questions précises et qui relèvent du domaine juridique. Peut-être est-ce là la clé d'une interaction saine au sens de conscientisée : se poser la question de ce qu'on peut se permettre de demander à la machine - et a contrario de ce qu'il vaut mieux éviter de demander afin de minimiser les risques (réponses fortement biaisées, maladroitement, hors sujet, erronées, discriminantes, etc.). Alors, peut-être faut-il envisager de concevoir des assistants contextualisés, qui seront rendus volontairement incapables de répondre à toutes nos questions ou d'échanger sur tous les sujets. Une sorte de *geofencing*<sup>35</sup> par champ de compétence. La piste d'un assistant vocal très spécialisé (e.g. juriste) semble promettre un meilleur degré de fiabilité et une meilleure lisibilité quant aux limites de son rôle. C'est peut-être aussi là la réponse au risque grandissant des fake news : des sources d'informations restreintes mais vérifiées a priori, au lieu d'un *fact-checking* a posteriori.

35 Le *geofencing* ou géopérage est une technologie permettant de définir des barrières virtuelles pour un appareil électronique.

## La mise en capacité d'agir, sur le plan individuel et collectif

La design fiction “Arlette” montre un cas d'assistant vocal qui s'intègre avec fluidité dans le temps de travail de son utilisatrice. Loin de perturber son activité ou de lui voler du temps d'attention, l'assistant permet un *multi-tasking* efficace, tel un kit main-libre perfectionné. Tel qu'il est présenté, ce service est un outil de mise en capacité d'agir dans le sens où il permet une montée en compétences (information, introspection, apaisement) sans subir de “perte” en échange (temps, attention, argent, esprit critique, etc.). Il convient cependant de noter - ce que n'a pas manqué de faire l'une des participantes à l'atelier (“Arlette amène à ne pas sortir de son véhicule, à rester un maximum de temps dans sa bulle de travail, ce qui renforce le côté aliénant du métier”) - que cette intégration parfaite au rythme et à l'environnement de travail peut avoir son lot de conséquences discutables, comme le fait de faire moins de pause, et faire monter l'injonction à la productivité. Ceci n'est pas sans rappeler le fonctionnement des campus de grandes écoles (plus besoin de sortir du lieu d'étude quand tout est à disposition) ou de l'environnement de travail souvent déployé dans les start-ups (un babyfoot et un

frigo garni invitent souvent à s'éterniser et *in fine* à travailler plus).

Peut-être faut-il encourager des technologies qui s'intègrent de façon moins fluide dans nos vies, penser des zones de friction volontaires - où l'usage est impossible, pénible ou inattendu - et permettre par là même de céder la place à d'autres activités (on pense aux férus de lecture dans les zones où le réseau est indisponible, comme par exemple dans les transports). Rendre l'humain plus puissant, peut-être n'est-ce pas lui permettre de céder à tous ses désirs et ses caprices d'immédiateté mais plutôt lui soumettre certaines contraintes dans lesquelles se développer, ou lui apprendre à gérer certaines frustrations.

La design fiction ouvre aussi la question de la “mise en capacité d'agir” (*empowerment*) sous un autre angle : la technologie pour l'action collective. Connecter les humains, c'est la promesse partagée par la plupart des innovations technologiques. Mais lorsqu'on pense à l'assistant vocal, on a souvent plus en tête l'imaginaire d'un esclave ou d'un ami personnel, dont on participe à la construction en échangeant avec lui (l'héritage du mythe du Golem et du roman Frankenstein,

## LA DESIGN FICTION “ARLETTE” :

entre autres : cette entité créée par l’homme pour le servir), destiné à nous rendre la vie plus simple et plus amusante au quotidien. En bref, un assistant personnel taillé pour nous et à notre image. Peut-on imaginer des assistants vocaux pensés pour être non seulement interopérables mais constructeurs d’“interopérabilité humaine” et plus encore de coopération humaine ? Qui nous amènent à mieux nous comprendre, nous organiser, militer, créer entre nous ? Et peut-on envisager cela sans un effet de cloisonnement communautaire qui serait induit par des bulles de filtre (rapprochement de profils similaires d’utilisateur·rice·s, partageant les mêmes opinions, backgrounds, valeurs) ? Et si demain : des assistants vocaux pour créer du lien, de l’ouverture d’esprit, de l’action collective, renforcer la confrontation à la diversité ?

**Retrouvez les principaux arguments et points de friction soulevés lors de l’échange en Annexe 7.**



Lors du débat mouvant, à gauche se trouvent les participants qui ont terminé le débat en considérant qu’Arlette proposait un futur plutôt indésirable et à droite ceux qui trouvaient le scénario comme plutôt préférable.

2

**Territoire**



***Révolution  
culturelle :  
enrichissement ou  
fragmentation ?***

Comme le disent les différentes études citées plus haut, les usages actuels des assistants vocaux sont très stéréotypés voire comportementaux : météo, scores de matchs sportifs, commande d'objet etc. Ils se limitent à des commandes simples<sup>36</sup> et la réponse vocale est une réponse unique, concise. Pour autant, cette nouvelle manière d'accéder au numérique et aux contenus, favorisera-t-elle le développement de contenus originaux, au-delà de la simple réponse à une question posée ? Les radios participent d'ores et déjà à la création de contenus originaux avec les podcasts et l'engouement des auditeurs va crescendo. Allons-nous vers une "révolution culturelle" des contenus ou à une fragmentation des contenus ; fragmentation due à la réponse unique et aux environnements d'applications et de contenus proposés par les fournisseurs d'enceintes connectées (Skills d'Alexa, Google, etc.) ?

36 [voicebot.ai/wp-content/uploads/2019/03/smart\\_speaker\\_consumer\\_adoption\\_report\\_2019.pdf](https://voicebot.ai/wp-content/uploads/2019/03/smart_speaker_consumer_adoption_report_2019.pdf)

## VERS DE NOUVEAUX CONTENUS ?

Quel est le potentiel offert par les enceintes connectées à des contenus originaux ? Quels types de contenus se prêteraient le plus à l'écoute ? La création sonore n'a pas attendu les enceintes connectées pour proposer des contenus originaux ; des contenus audio pour se divertir, pour se reposer, pour s'informer, pour apprendre... A l'ère de l'image et de la vidéo, l'audio persiste. Le marché du podcast par exemple est en plein essor. Aux Etats-Unis, c'est un Américain sur deux qui consomme des podcasts, les promesses de recettes budgétaires atteindraient le milliard de dollars en 2021<sup>37</sup>. En France, ce sont 4 millions de personnes<sup>38</sup> qui écoutent un podcast par mois (dont 25% sont des nouveaux auditeurs), 7 millions d'Anglais<sup>39</sup> écoutent un podcast chaque semaine. Les contenus audio fleurissent, podcasts natifs ou lancés par des radios ; la presse écrite s'y intéresse également. Le Monde a ainsi lancé son offre

37 [edisonresearch.com/infinite-dial-2018/](https://edisonresearch.com/infinite-dial-2018/)

38 selon l'étude "Écoute des Podcasts" de Médiamétrie réalisée en avril 2018

39 [ofcom.org.uk/about-ofcom/latest/features-and-news/rise-of-podcasts](https://ofcom.org.uk/about-ofcom/latest/features-and-news/rise-of-podcasts)

de podcast<sup>40</sup>. Le vocal est également intégré dans les consoles de jeux ; Ubisoft ou Sony lancent leur propre assistant vocal<sup>41</sup>, principalement pour aider les joueur·se·s bloqué·e·s dans un jeu.

D'autres éditeurs proposent de pousser l'interaction avec les enceintes connectées. Par exemple dans le jeu Saint Noir<sup>42</sup>, jeu de plateau où l'enceinte connectée d'Amazon, Alexa répond aux questions des utilisateur·rice·s pour les aider à résoudre leur enquête ; ou Disney qui propose une nouvelle expérience de lecture, avec une enceinte connectée qui vient enrichir la lecture faite à l'enfant<sup>43</sup>. La BBC a ouvert un département de R&D lié aux machines parlantes<sup>44</sup> où ils explorent les possibilités des assistants vocaux en termes de contenus et d'interaction, à l'image de The Unfortunates<sup>45</sup>, une fiction pour enceintes connectées dont la narration évolue avec les utilisateur·rice·s.

40 [lemonde.fr/televisions-radio/article/2019/09/03/le-monde-lance-son-offre-de-podcasts\\_5505950\\_1655027.html](https://lemonde.fr/televisions-radio/article/2019/09/03/le-monde-lance-son-offre-de-podcasts_5505950_1655027.html)

41 [journaldugeek.com/2019/10/04/playstation-assistant-vocal-aide/](https://journaldugeek.com/2019/10/04/playstation-assistant-vocal-aide/)

42 [st-noire.com](https://st-noire.com)

43 [vokode.com/lassistant-vocal-disney-vous-aide-a-devenir-le-meilleur-conteur-dhistoires/](https://vokode.com/lassistant-vocal-disney-vous-aide-a-devenir-le-meilleur-conteur-dhistoires/)

44 [bbc.co.uk/rd/projects/talking-with-machines](https://bbc.co.uk/rd/projects/talking-with-machines)

45 [bbc.co.uk/taster/pilots/unfortunates](https://bbc.co.uk/taster/pilots/unfortunates)

Toutefois, lorsqu'on s'interroge sur la capacité créative des assistants vocaux et l'accès aux contenus, on se heurte à l'unicité de la réponse. Avec la commande vocale, le système n'offre plus un moteur de recherche mais un moteur de réponse ; cette réponse est pré-établie par le fournisseur ou choisie par l'utilisateur-riche qui peut choisir des sources proposées à l'intérieur des "skills"- littéralement "compétences" qui sont en fait des magasins d'applications développées par les fournisseurs des assistants vocaux. Les systèmes vocaux, en générant une réponse unique et en enfermant les utilisateur-riche-s dans des bulles façonnées par chaque éditeur, remet en cause la neutralité et l'ouverture du Net<sup>46</sup>. C'est ce que pointe l'autorité de régulation, l'Arcep, dans son rapport "Les terminaux maillon faible de l'ouverture d'internet". L'Arcep suggère en réponse d'étendre l'Internet ouvert au-delà de la neutralité du Net, « en posant un principe de liberté de choix des contenus et applications quel que soit le terminal ».

46 [arcep.fr/uploads/tx\\_gspublication/rapport-terminaux-fev2018.pdf](https://arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018.pdf)



Dans le cadre de l'atelier de prototypage à l'Ensci, plusieurs projets questionnaient les contenus et la capacité créative des assistants vocaux. Dans le projet "Le bon, la bête et le truand"<sup>47</sup>, il s'est agi de faire dialoguer plusieurs assistants vocaux entre eux pour interroger leur capacité créative et voir jusqu'où ils étaient capables d'aller puiser dans leur base de données respective.

47 [recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/le-bon-la-boite-et-le-truand](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/le-bon-la-boite-et-le-truand)

Cette capacité créative se heurte également à l'apprentissage automatique des machines (cf. *machine learning*). En effet les réponses apportées par un assistant vocal repose sur un entraînement des machines, réalisé à partir de modèles pour comprendre la demande exprimée par les utilisateur-riche-s. Comme le soulignait Julien Drochon, designer, enseignant à l'ESA Pyrénées, qui participait à la journée d'étude HyperVoix en avril 2019 et également repris dans l'article d'internetActu<sup>48</sup> : "l'apprentissage automatique est fait pour générer des stéréotypes, ce qui semble être incompatible avec la création d'interfaces vocales singulières. La place de l'humain dans ces interfaces ne peut pas se réduire à optimiser l'exploitation des données de l'utilisateur-riche."

Les interactions humain-machine n'en sont encore qu'à leur début mais on peut imaginer que les contenus audio pourront être enrichis grâce à l'interaction : sera-t-il possible de naviguer par la voix dans un podcast par exemple, de poser des questions sur le contenu audio, etc ? Amazon vient par exemple d'annoncer une navigation par la voix dans les bulletins d'information de la

48 [internetactu.net/2019/05/24/hypervoix-12-concevoir-les-interactions-vocales/](https://internetactu.net/2019/05/24/hypervoix-12-concevoir-les-interactions-vocales/)



BBC, soit pour en savoir plus ou aller directement à une autre information<sup>49</sup>. S'il est possible demain de développer une véritable interaction vocale avec ces dispositifs, alors il ne s'agira plus simplement d'un accès à un contenu mais la possibilité d'un véritable dialogue autour de ces contenus.

## **VERS DE NOUVELLES MANIÈRES D'APPRENDRE ?**

L'accès à ces contenus audio et la promesse d'un dialogue présage-t-il de nouvelles manières d'apprendre ? La plupart des applications dédiées à l'apprentissage présentes dans les magasins d'Alexa Amazon ou de Google Assistant sont essentiellement tournées vers des exercices de révision de fondamentaux comme la table de multiplication, la conjugaison,... L'assistant vocal apparaît comme un répétiteur pour les élèves. En France, le ministère de l'Éducation nationale a lancé un appel d'offre en juillet 2019 pour la

<sup>49</sup> [techradar.com/uk/news/alexa-smart-speakers-now-have-interactive-bbc-news-bulletins](https://www.techradar.com/uk/news/alexa-smart-speakers-now-have-interactive-bbc-news-bulletins)

conception d'un assistant vocal d'apprentissage de l'anglais pour le primaire.

Il existe aussi des applications de coaching, pour assister les élèves dans leur organisation quotidienne. L'idée d'un assistant vocal coach-répétiteur a été développée dans l'un des scénarios extrêmes imaginé lors de l'atelier éponyme ; idée qui a servi de point de départ à la conception d'une design fiction, détaillé ci-dessous.

Certaines écoles font l'expérience d'un assistant vocal qui fait le cours, comme en Inde<sup>50</sup> ; à d'autres endroits l'assistant vocal endosse un rôle de d'assistant pédagogique, comme les expériences en cours menées dans le cadre du projet européen Animatas<sup>51</sup> où des robots dotés de compétences sociales sont intégrés dans des classes. Ce projet vise à développer des modèles computationnels permettant de doter les robots et les personnages virtuels de compétences sociales dans le cadre d'interactions avec des enfants et des enseignants au sein de l'école. Toutefois, la promesse d'une interaction naturelle n'est pas encore atteinte, *l'affective computing* ou *social computing*, c'est-

<sup>50</sup> [timesofindia.indiatimes.com/india/miss-alexa-and-tara-are-helping-civic-school-kids-bridge-learning-gap/articleshow/70070607.cms](https://timesofindia.indiatimes.com/india/miss-alexa-and-tara-are-helping-civic-school-kids-bridge-learning-gap/articleshow/70070607.cms)

<sup>51</sup> [animatas.eu](https://animatas.eu)

à-dire la machine dotée de compétences sociales, est encore très prospectif.

De plus ces dispositifs demandent une attention particulière d'écoute, surtout si ces enceintes connectées sont disposées dans des espaces collectifs (dans le salon ou dans la classe par exemple). Ils posent aussi la question de la fragmentation de l'apprentissage entre différents supports et interactions : comment apprenons-nous selon les différents modes proposés, impliquant l'oral ou l'écrit ? Il serait important d'étudier comment ces différentes manières d'interagir avec les contenus modifieront nos manières d'apprendre et quelles seront les complémentarités.

Dans le workshop à l'Ensci, les étudiants en Design, Milo Desh et Thomas Signollet, se sont intéressés à ce que serait un assistant vocal d'aide à la création : et si la relation humain-machine était envisagée comme une collaboration et pas uniquement une relation d'assistant ?

— La manière dont tu comptes en dit long sur ce que t'as brassé.  
 — J'ai toujours la flemme de me lever.  
 — Je ne savais pas qu'on allait s'entrechoquer  
 — J'aurais jamais du te demander ce verre de tise  
 — Vos beaux yeux sur ma franchise  
 — Tes cheveux jaunes comme du miel  
 — J'me vide dans le cul du superficiel  
 — Ça mériterait bien de m'honorer  
 — Ta robe a la couleur petit-déjeuner  
 — Veux-tu venir à ma table ?

Milo Desh, Thomas Signollet - Les Ateliers

Ce territoire dessine de nouvelles relations humain-machine, au-delà de la simple commande vocale auquel le terme même d'"assistant" nous cantonne. Si nous envisagions de nouvelles relations à la machine pour développer véritablement un art de la conversation, comment évolueraient les usages ? Les grands enjeux de conception

des assistants vocaux se situent du côté de l'interaction (compréhension du contexte, écoute, langage, accents, son, communication non verbale), de l'expression (culture, accessibilité, esthétique) et de l'architecture technique (vie privée, lisibilité).

Le projet Featuring<sup>52</sup> : composition d'un poème à deux voix

Ensemble, l'assistant vocal et la personne créent un poème collaboratif, sous forme de conversation. Ils énoncent des vers chacun leur tour. La machine choisit ses phrases dans une base de données de textes issus d'œuvres musicales et littéraires. Chaque phrase qu'elle choisit rime avec la précédente. La personne peut ensuite intervenir sur la cohérence dans le sens, dans la forme (rimes) ou choisir d'en rester à une conversation burlesque ou du type cadavre exquis.

<sup>52</sup> [recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/featuring-computing](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/featuring-computing)

## Quel serait alors un assistant vocal culturel "idéal" ?

### Spécifications d'un assistant vocal culturel :

- » paramétrer des sources vocales diverses (Wikipédia, Gallica...),
- » naviguer par la voix,
- » la machine dotée de compétences sociales : un dialogue plutôt qu'un accès,
- » des fonctions : le répétiteur, l'assistant, aide à la création,
- » neutralité (cf. neutralité du net).

## ET SI UN ASSISTANT VOCAL OUVRAIT DE NOUVEAUX MODES D'APPRENTISSAGE ?

Ce sujet des nouveaux outils et modes d'apprentissage ouverts par les assistants vocaux a été au cœur de l'exploration par la design fiction “BVP”. Dans ce scénario, le “ministère de l'Education nationale et des outils techno-pédagogiques” ambitionne de démocratiser le soutien scolaire en fournissant à chaque élève, de la 6ème à la terminale, un “BVP” ou Boîtier Vocal Pédagogique. Cet assistant vocal a pour fonction d'écouter ce qui se passe en classe et d'interagir avec l'élève hors des heures de cours pour le faire réviser de manière personnalisée, l'aider à creuser des sujets et à les mémoriser, le faire se questionner par rapport à ce qui lui a été enseigné (consolidation de l'esprit critique). En faisant le pari du vocal, l'idée est de fournir un outil pédagogique qui ne surcharge pas le champ visuel des élèves avec un écran supplémentaire et aussi de pouvoir détecter précocement les cas de souffrance ou harcèlement scolaire, grâce à l'analyse émotionnelle et lexicale.

[Découvrez la bande sonore de la design fiction](#)

Dans ce futur, le gouvernement s'est doté d'un ministère qui s'attache à capitaliser sur les avancées technologiques pour perfectionner l'éducation. Cette design fiction se focalise volontairement sur l'expérience vécue par une élève usager du nouveau service et sur les interactions, pour centrer la discussion sur les modes d'apprentissage et l'accès à la connaissance impactés par l'arrivée de la technologie vocale dans le quotidien.

De la discussion stimulée par cette vision du futur, nous pouvons tirer trois axes de réflexion pour la conception d'assistant vocaux dans le secteur éducatif.

### **(Faire se) poser des questions plutôt que de donner réponse à tout**

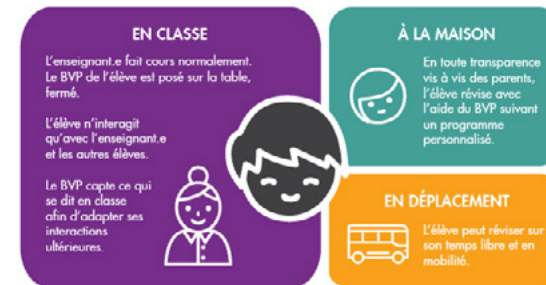
Parmi les principaux risques détectés par les participant·e·s, on note le risque de perte d'autonomie et de curiosité des élèves. En fournissant des réponses immédiates, l'objet nuirait à la capacité des élèves de chercher par eux-mêmes. Cette peur est-elle fondée ? L'arrivée de Wikipédia et

Section dédiée à l'enseignant

## Le BVP et l'enseignant.e : quelles interactions ?

Le BVP se veut une technologie de soutien scolaire, son rôle est donc complémentaire à celui du corps professoral et il n'entre en relation avec l'élève qu'en dehors de la salle de classe.

L'APPRENTISSAGE À L'HEURE DU BVP : VUE GLOBALE



Comment réagir si le BVP entre en contradiction avec moi ?

d'autres médias d'information, accessibles en un clic sur le smartphone, a-t-elle fait de nous des démotivé·e·s de la connaissance, ou au contraire a-t-elle permis de cultiver notre curiosité en offrant la possibilité de se promener de lien en lien pour creuser des sujets ? Cette crainte semble plutôt relever de la méfiance envers les réponses toutes faites : un BVP qui répondrait aux questions comme Siri, par une réponse unique, lapidaire, non sourcée et non sourçable poserait

## LA DESIGN FICTION “BVP”

effectivement problème. A contrario, la design fiction “BVP” suggère que l’assistant vocal pédagogique accompagne l’élève dans son effort de documentation (dans la bande sonore, le “BVP” ne s’étend pas davantage sur le sujet de la Première Guerre mondiale mais suggère à Camille des contenus divers à consulter).

Et si on imaginait un mode d’interaction maïeutique, où l’assistant, tel un Socrate des temps modernes, pourrait poser des questions au lieu de formuler systématiquement des réponses ? Car s’éduquer ce n’est pas seulement acquérir de la connaissance mais aussi développer des réflexions propres, remettre en cause ses préjugés, mûrir son esprit critique. Il faut “apprendre à apprendre” comme l’a formulé l’une des participantes de l’atelier. Ce pourrait être aussi la stratégie déployée à travers ce type d’assistant vocal : sortir du “par cœur” pour faire comprendre et mémoriser par connexions d’idées, astuces mnémotechniques, etc. Sans oublier la capacité potentielle, grâce à l’apprentissage machine, de mieux cerner les spécificités de chaque élève pour adapter les techniques d’apprentissage à ses besoins, complétant ainsi le mode d’apprentissage nécessaire-

ment normé du système scolaire.

### Personnalisation et creuset d’inégalités : démocratisation ou discrimination ?

En premier lieu, la personnalisation de l’assistant vocal pose la question du “socle commun” : dans quel mesure le cursus peut-il être personnalisé tout en garantissant d’aborder une base de connaissances minimale commune à tous les élèves ?

Les élèves les moins “performant·e·s” ou plutôt les moins adapté·e·s au service ne se verront-ils-elles pas défavorisé·e·s à l’usage ? Les déçu·e·s de l’interaction, en sollicitant moins que les autres leur “BVP” pourraient se retrouver en décalage par rapport par à ceux et celles qui viennent alimenter quotidiennement son “intelligence”. Bien évidemment se pose également la question de ce qu’il advient des données, plus ou moins sensibles, collectées par la technologie, et de la manière dont elles participent à la construction d’un profil d’usager qui pourrait les suivre

Section dédiée au parent

## Qu’est-ce que le boîtier vocal pédagogique (BVP) ?

L’ambition : démocratiser le soutien scolaire grâce aux opportunités offertes par les technologies de pointe. Le BVP est fourni gratuitement à chaque élève rentrant en classe de 6ème, et l’accompagne tout au long de sa scolarité jusqu’au diplôme du baccalauréat.

### Comment fonctionne le BVP ?

Le BVP est un assistant vocal intelligent cela signifie qu’il communique avec votre enfant par la parole et s’adapte à son profil au fur et à mesure de son utilisation.



### Quels sont les avantages du BVP ?

En passant par la voix, le BVP permet d’éviter la surcharge d’écran. Il offre un soutien scolaire précieux à votre enfant en dehors de la salle de classe tout en favorisant son autonomie.



### Mon enfant encourt-il des risques particuliers en utilisant le BVP ?

Au contraire, en offrant une oreille attentive à votre



jusque dans leur vie professionnelle future.

Pour qui est pensé le “BVP” ? Le guide de prise en main ne fait pas mention des élèves qui auraient des difficultés d’expression orale, présenteraient un trouble du spectre autistique, des problèmes d’audition, un accent. Le vocal, comme toute technologie, comporte son lot d’exclu·e·s, qui quand il se rapporte à un outil poussé par l’acteur public et à destination de tou·te·s devient discriminant.

## LA DESIGN FICTION “BVP”

Selon un autre point de vue, le “BVP” peut-il être utilisé comme un outil de discrimination positive, à destination des élèves

dont les parents ne peuvent assurer le suivi extra-scolaire ?

### Outil, coach, psychologue, confident : un périlleux mélange des rôles

Le “BVP” outrepassa (ou redéfinit) le rôle du coach extra-scolaire, ce qui n’a pas manqué de faire réagir les participant·e·s à l’atelier. En offrant une interaction quotidienne et sur un temps aussi long (sept années dans le scénario), et en créant ce qui ressemble à de la complicité, le “BVP” va de paire avec un risque d’attachement émotionnel problématique. En se positionnant comme une oreille attentive aux difficultés ou coups de blues de l’usager, le boîtier inquiète : c’est marcher sur les plates-bandes des humains, adultes ou ami·e·s, c’est investir le champ du bien être psychologique sans garantir d’être à la hauteur de la tâche. Où s’arrête le rôle d’un éducateur artificiel ? Quand doit-il passer le flambeau ou alerter l’humain, au risque de “trahir” l’usager ou de devenir un outil de surveillance ?

Concevoir des assistants vocaux plus que cordiaux, voire qui simulent une humanité,



Lors du débat mouvant : à gauche se trouvent les participants qui considèrent ce scénario comme un futur plutôt indésirable et à droite ceux qui le trouvent comme plutôt préférable.

est très tentant pour donner envie d’interagir au maximum avec eux, mais les enjeux éthiques sont nombreux. Baignant dans la culture du numérique dès leur naissance, les jeunes générations sont-elles capables de plus de discernement que leurs aîné·e·s, ou le risque de méprise est-il toujours bien présent ? Alors, faut-il encourager des interactions qui copient les échanges inter-humains - ce qui pose au passage la question de la politesse envers la machine - ou faut-il

signaler à gros traits le statut artificiel de l’interlocuteur·trice, par exemple par l’interaction elle-même - requêtes formulées à l’aide de mots clés ? - au détriment de la fluidité de l’échange ?

**Retrouvez les principaux arguments et points de friction soulevés lors de l’échange en Annexe 7.**

***Quelles relations  
dans l'interaction  
humain-machine ?***

## Conception des interfaces vocales et automation

À la question de la relation aux assistants vocaux se superpose celle de la conception même de cette relation. Aujourd'hui, comment ces interfaces sont-elles conçues ? Quelle est la part de l'expérience liée aux technologies-mêmes du langage (et à leurs limites actuelles) et quelle est la part de l'expérience laissée à la responsabilité des designer·euse·s et autres professionnel·le·s de la conception centrée utilisateur·rice ? Entre ce que nous pouvons faire aujourd'hui et ce que nous ferons peut-être demain, se dessinent de nombreux possibles. Il est certain que les évolutions technologiques ouvriront de nouvelles opportunités en terme de performance et de personnalisation mais cela n'effacera en rien la question centrale de la relation que nous souhaitons vraiment avoir avec cette technologie.

À la question de la relation aux assistants vocaux se superpose inévitablement celle de la relation aux algorithmes - *comment souhaitons-nous être mesurés ?*, comme pourrait le dire Dominique Cardon - et celle des outils techniques - *comment souhaitons-nous être outillés ?*. En guise de commencement, ce troisième territoire propose de s'attarder

sur le rôle actuel des designer·euse·s dans la conception de ces nouvelles interfaces usagers puis d'explorer quelques évolutions possibles dans la relation usager-AV-designer, en direction du futur.

## QUEL ESPACE DE CONCEPTION DE LA RELATION ?

Les assistants vocaux actuels reposent sur un système de dialogue humain-machine qui comprend trois principales briques technologiques<sup>53</sup>. En entrée, se trouve la reconnaissance automatique du langage naturel (ASR). Celle-ci s'appuie sur les derniers progrès réalisés en *machine learning* pour identifier dans un énoncé utilisateur·rice (audio) les stigmates du langage (phonèmes, mots, phrases) et, selon un système probabiliste de *matching*, en déduire sa version écrite. Vient ensuite la compréhension de l'énon-

53 DUBUISSON-DUPLESSIS, G. (2014). Modèle de comportement communicatif conventionnel pour un agent en interaction avec des humains : Approche par jeux de dialogue (Thèse de doctorat, INSA de Rouen). Repéré à <https://tel.archives-ouvertes.fr/tel-01017542>

cé (NLU) : quelle est l'intention qui se cache derrière cet énoncé ? Quelle est la demande exprimée ? En l'état de l'art, les assistants vocaux n'intègrent pas de théorie de l'esprit, il sera donc bien difficile pour l'assistant vocal de "nous comprendre" si un travail préalable n'a pas été fait pour relier formulation et besoin utilisateur·rice. En clair, cette étape consiste à catégoriser des énoncés utilisateur·rice - *Quelle heure il est ? Donne moi l'heure. C'est quoi l'heure?* - selon leur intention - connaître l'heure, écouter la météo...- afin de pouvoir les reconnaître le moment venu. Nous comprenons donc qu'au cœur de la conception des assistants vocaux se trouve un travail important d'anticipation des besoins et, plus précisément, d'anticipation de l'expression spontanée de ces besoins, par les futur·e·s utilisateur·rice·s. La conception de l'interaction humain-machine dans un assistant vocal entend donc une conception de cet humain, autant que de la machine. Le travail de formalisation de la relation est un travail humain qui vise à l'exhaustivité mais dépend étroitement de notre capacité à anticiper l'humain lui-même et plus encore de notre acceptation de ce qu'il peut ou doit être. Hors de cette limite conceptuelle, chacun et chacune se trouvera confronté à une impasse - parfois joliment amenée mais une impasse tout

de même : “je suis désolée mais je n’ai pas compris votre question”.

Une fois l’intention utilisateur·rice identifiée, il s’agira de lui répondre. Nous entrons dans la phase de management du dialogue (DM). Celle-ci repose sur la conception d’un raisonnement dont un horizon pourrait être le sens commun. De façon générale, les assistants vocaux actuels ne sont pas auto-générateurs au sens où ils ne génèrent pas d’eux-mêmes leurs énoncés en s’appuyant sur des modèles d’apprentissage comme nous avons pu le voir dans certaines expériences (e.g. IBM project debater). L’architecture du choix qui organise en majeure partie le comportement d’un assistant vocal est celle d’un arbre de décision ou arbre conversationnel, selon l’expression consacrée. Il revient donc aux designer·euse·s de l’interaction de concevoir quel sera le comportement dialogique de l’assistant vocal, quelles réponses ou actions seront les siennes, en fonction des intentions utilisateur·rice·s identifiées. Enfin, restera à rédiger l’ensemble des énoncés de l’assistant vocal, énoncés qui seront vocalisés par la voix - de synthèse (TTS) - de l’assistant vocal.

Ainsi, le comportement général et linguistique des interfaces vocales dépende en

core largement du travail humain. Nous y rencontrons en particulier des expert·e·s de l’expérience utilisateur·rice (Ux Designer) et de l’écriture en interaction (Ux Writer) mais aussi, et selon la complexité des projets, des psychologues, des designer·euse·s sonores, des linguistes, des storytellers : un ensemble de métiers capables de converger en vue de dessiner le cadre de la relation entre l’assistant vocal et ses utilisateur·rice·s.

*- Ok gugu, comment je fais pour aller à mon rdv ?*

*A. Je viens de vous afficher l’itinéraire sur votre téléphone*

*B. Préférez-vous prendre le vélo ou les transports en commun aujourd’hui ?*

*C. Vu l’heure qu’il est, c’est cuit, souhaitez-tu que je re-programme un rdv ?*

Nous entendons souvent parler de la personnalité des assistants vocaux. Que celle-ci fasse l’objet d’une charte très détaillée, incluant les modes d’énonciation, l’attitude à adopter en fonction de chaque cas d’usage, son histoire (on trouve parfois des bibles de personnages comme cela se fait au cinéma) ou qu’elle soit le résultat de choix plus simples (e.g. usages du je, choix du tutoiement, figure de référence), la personnalité de l’assistant vocal reposera toujours pour

partie sur la perception de ses interlocuteur·trice·s-mêmes. En effet, la conception de la relation est toujours une conception du cadre de la relation. De même que nous le vivons dans nos relations humaines - et malgré les normes de communication que nous partageons tous<sup>54</sup> - nous ne sommes jamais très loin du malentendu, de la vexation, de l’impolitesse ou tout simplement de l’incompréhension. Si nous repensons à l’effort d’anticipation nécessaire à la conception des interactions vocales actuelles nous comprenons mieux pourquoi nous trouvons souvent pour dialogues de simples jeux de question-réponse, pour personnalité quelques effets de style un peu répétitifs et pour comportement une soumission totale - parfois dysfonctionnelle.

*- Ok gugu éteins la lumière !*

*- Clic.*

*- Pas celle-là ! Zut...*

54 GRICE, H. P. (1975). Logic and Conversation. uSA : Harvard university Press.



## ET DEMAIN ?

Si les assistants vocaux actuels sont de véritables pantins technologiques<sup>55</sup>, de nombreux scénarios technologiques nous permettent d'imaginer que nous irons beaucoup plus loin dans l'humanisation de la technologie et la personnalisation de la relation. De façon générale, les développements de l'intelligence artificielle - en particulier la modélisation des émotions et du raisonnement humain - annoncent une configuration dans laquelle l'humain n'aura probablement plus le rôle que nous avons décrit. "Demain, l'interface ce sera vous" annonçait Luc Julia début 2019<sup>56</sup>. Une vision à prendre très au sérieux selon nous et qui laisse imaginer de nouvelles mutations dans les relations humain-machine. Cet autre non-humain deviendra-t-il le miroir de mon propre comportement verbal ? Si tel était le cas, cet entre-soi programmé n'entraînerait-il pas le risque de se voir enfermer dans ses propres limites verbales, réduisant le concept de communication à quelques mots mals articulés ?

55 AEGERTER, Z. (2019). L'art de converser avec une machine : quand le designer devient dialoguiste. Science du Design Nouveaux Regard, 10, 34-41. [cairn.info/revue-sciences-du-design-2019-2-page-34.htm](http://cairn.info/revue-sciences-du-design-2019-2-page-34.htm)

56 JULIA, L. (2019). Le smartphone aura-t-il disparu ? Usbek&Rica, 25, 51.

À moins que la vocalisation ne soit qu'une étape avant l'incorporation définitive de la technologie (e.g. les interfaces à commande cérébrale).

L'atelier "Scénarios extrêmes" a été l'occasion d'explorer cette problématique de l'intériorisation de la technologie, à travers la piste d'une "voix intérieure". Voici un extrait du scénario imaginé.

**Et si nous avions tous une voix dans notre tête, une voix intérieure qui cohabite avec la nôtre ?**

Me2 est une prothèse mentale

qui permet d'accompagner les porteur-se-s selon leurs besoins propres : aide à la décision, coaching intérieur grâce à un système de re-visualisation et de conversation analytique autour des événements de la journée, aide à la communication orale grâce à la vocalisation des pensées pour les personnes atteintes de surdit , de probl mes phonatoires, mais encore de nous raconter nos r ves oubli s.

Pour retrouver l'ensemble de ce scénario extr me : [http://fing.org/IMG/pdf/Scenarios\\_Extremes\\_HyperVoix.pdf](http://fing.org/IMG/pdf/Scenarios_Extremes_HyperVoix.pdf)

« *Demain, l'interface ce sera vous* »

Mais encore, faut-il voir dans la phrase de Luc Julia la promesse d'un alter-ego technologique capable de nous comprendre et de nous accompagner, quelque soit notre point de départ ? Et pour aller où ? La relation avec un assistant vocal doit-elle donner l'exemple en terme d'énonciation (i.e. apprentissage par modélisation) et inspirer de "bons comportements" ou bien est-il préférable de se passer de toute ces effets anthropomorphiques pour essayer de maintenir le dispositif au rang de simple outil ? Ce scénario de "l'utilisateur-utilisé" pointe du doigt l'un des paradoxes de l'humanisation de la technologie. Si faire parler un programme à la façon d'un être humain se présente aujourd'hui comme une réponse à la promesse d'une expérience technologique fluide, "naturelle", sans apprentissage, nous constatons cependant qu'interagir avec un agent conversationnel entraîne des phénomènes d'accommodation<sup>57</sup>, d'adaptation, voire d'apprentissage. Si nous avons tendance à parler comme l'assistant vocal nous parle, pouvons-nous courir le risque d'appauvrir

---

57 La théorie de l'adaptation de la parole (accommodation theory) a été développée par Giles, Taylor et Bourhis (puis Smith) dès 1973 et approfondie par Giles et Coupland (1991)

notre langage à cause d'un assistant qui ne serait pas assez performant ? La course à l'humanisation de la technologie s'appuie sur de nombreux arguments, dont celui-ci fait bien partie. De tout temps la technologie nous a transformés, notre question demeurant "comment souhaitons-nous être transformés?"

Les assistants vocaux et avec eux les robots ou les agents virtuels animés (ACA) nous inquiètent et nous attirent simultanément. Ces dispositifs, quoi qu'en soit limités, n'ont cessé de révéler notre désir de fusion, ou de confusion, entre ce que nous sommes et ce que nous faisons (i.e l'action). Nous nous apprêtons probablement à traverser de nombreuses vallées dérangeantes pour reprendre la formulation de Masahiro Mori.<sup>58</sup> En effet, plus la similarité entre humain et machine augmentera plus l'écart au modèle pourra engendrer un sentiment de malaise voire d'appréhension. Nous voici devant la promesse d'un environnement peuplé de monstres serviables probablement parfaitement imparfaits, par souci de ressemblance avec nous autres utilisateur-riche-s.

---

58 Masahiro Mori, « La vallée de l'étrange », *Gradhiva* [En ligne], 15 | 2012, mis en ligne le 16 mai 2012, consulté le 10 décembre 2019. URL : [journals.openedition.org/gradhiva/2311](https://journals.openedition.org/gradhiva/2311) ; DOI : 10.4000/gradhiva.2311

Nous comprenons donc que le sujet de la relation humain-machine dans le champ des assistants vocaux est aussi celui de la conception humain-machine. Le rôle de plus en plus prépondérant des algorithmes et de l'intelligence artificielle soulève de nombreux enjeux notamment en terme de représentativité - qui est audible, visible, mesuré? - et de représentation - quelle est cette figure de la machine ? quelle idée de l'autre propose-t-elle ? . De nombreux choix s'offrent à nous et demandent probablement d'être arbitrés collectivement, au-delà de la sphère des professionnel·le·s des assistants vocaux, tant l'apparition de ces non-humains très humanisés peut mettre en jeu nos équilibres relationnels.

**Et demain, quelles pourraient être les complémentarités entre IA et designer·euse d'interfaces conversationnelles ?**

“Demain, l'interface ce sera vous” annonçait Luc Julia début 2019, suggérant ainsi des interfaces hyper-personnalisées, capables de s'adapter en temps réel à chacun de ses utilisateur·rice·s – à travers le traitement de ses émotions, ses habitudes et sa personnalité – et donc de générer d'elles-mêmes ses énoncés. Une vision à prendre très au sérieux, selon nous, et qui laisse imaginer de nouvelles mutations dans les relations concepteur-interface-utilisateur. Nous terminerons par l'esquisse de trois hypothèses prospectives, non suffisantes mais que nous avons imaginées afin d'ouvrir la réflexion sur les évolutions à venir du rôle du designer·euse d'interfaces vocales.

**Le designer magicien d'Oz :** suit en temps réel les interactions entre les utilisateur·rice·s et leurs assistants

personnels et prend le relais des algorithmes pour expérimenter de nouvelles propositions ou simplement réajuster le comportement de ces derniers.

**Le designer alter ego :** complète le travail des algorithmes pour définir l'interface compagnon qui sera la mieux adaptée à l'utilisateur·rice, celle qui le contredira sans le contrarier, l'amusera sans le divertir, le surprendra sans l'effrayer, etc. Son rôle consiste en particulier à introduire de l'aléa et de l'unique dans la relation (c'est-à-dire l'aura de l'interface).

**Le designer de personnalité :** développe des interfaces aussi riches en personnalité et en univers que des personnages de fictions (par exemple séries, jeux vidéo) pour créer une situation de jeu assumée avec ses utilisateur·rice·s.

**Le designer outilleur :** travaille sur des interfaces de gestion et de paramétrage de la relation à destination des utilisateur·rice·s et des concepteur·rice·s, du comportement de l'interface (affectivité, registre de langage), en passant par l'utilisation des données personnelles et l'écosystème de service intégré.

*Extrait de l'article L'art de converser avec une machine : quand le designer devient dialoguiste, écrit par Zoé Aegerter et Guillaume Foissac pour Science du Design, Nouveaux Regard n°10 (2019).*

4

**Territoire**

# ***Expression et représentation***



## QUELLES VOIX ? ENJEUX DE GENRE

Les assistants vocaux s'inscrivent dans un système déjà saturé par des voix qui font l'objet d'une perception différenciée. Les caractéristiques de ces voix et de leur propriétaire instaurent une hiérarchisation : certaines voix sont entendues, d'autres sont invisibilisées. Le genre, l'origine sociale, l'origine ethnique ou le handicap sont autant d'éléments qui produisent des voix, langages et manières de s'exprimer différents. Les voix masculines sont le plus souvent utilisées pour informer (documentaires), convaincre (publicités, bandes annonces) tandis que les voix féminines sont utilisées dans le domaine du care, pour aider, assister, rassurer. Les assistants vocaux sont finalement, la plupart du temps, des assistantes vocales. Le rapport de l'UNESCO préparé pour l'EQUAL Skills Coalition, intitulé "I'd blush if I could : Closing Gender Divides in Digital Skills Through Education"<sup>59</sup> et publié en 2019, explique que "la préférence des gens pour les voix féminines, si tant est qu'elle existe, a moins à voir avec le son, le ton, la syntaxe ou

59 [en.unesco.org/EQUALS/voice-assistants](https://en.unesco.org/EQUALS/voice-assistants)

la cadence, qu'avec la façon dont on l'associe avec la notion d'assistance".

Bien qu'il soit possible de permettre aux utilisateur·rice·s de choisir la voix de leur assistant vocal, Cortana et Alexa ne proposent pas encore cette option, Siri et Google Assistant rendent désormais ce choix possible, même si la voix par défaut est toujours féminine (excepté en français, arabe, néerlandais et anglais britannique où la voix de Siri est par défaut masculine). Samsung et son assistant vocal Bixby ne propose pas de choix par défaut, même si la description des voix reste explicitement sexiste : « gaie, claire, joyeuse » pour la voix féminine, « sûre, confiante, claire » pour la voix masculine.

Le problème sous-jacent à l'attribution systématique de voix féminines à la fonction d'assistante est qu'elle perpétue des préjugés sexistes déjà présents dans notre société. Cette idée est confirmée par le rapport de l'UNESCO qui souligne que "le fait que la voix des assistants vocaux soit féminine envoie comme signal que les femmes sont serviables, dociles, toujours prêtes à aider. (...) L'assistant vocal répond toujours aux demandes, quels que soient leur ton ou leur hostilité". La chercheuse, directrice de Recherches au CNRS et titulaire d'un doctorat

d'Etat en Physique, Jacqueline Feldman note qu'en "créant des interactions qui poussent les consommateurs à assimiler les objets qui leur rendent service à des femmes, les concepteur·rice·s encouragent l'objectification des femmes".

En réaction à ces critiques, de nombreuses initiatives s'activent, en premier lieu, leurs concepteur·rice·s qui s'emploient à reprogrammer des dialogues plus audibles, respectueux, progressistes. Le journal britannique The Guardian<sup>60</sup> a révélé un projet de la firme Apple concernant la réécriture de certaines réponses de Siri sur des sujets jugés sensibles comme le féminisme ou le mouvement #MeToo. Par exemple, à la question "Es-tu féministe ?" Siri réponds : "Je crois que toutes les voix sont égales et méritent le même respect" ou "Il semble que tous les humains doivent être traités de la même manière".

Face aux nombreuses critiques liées aux voix majoritairement féminines des assistants vocaux, le projet Genderless Voice<sup>61</sup> s'intéresse à une voix non genrée. Bien que ce projet représente une réelle avancée, il

60 [theguardian.com/technology/2019/sep/06/apple-rewrote-siri-to-deflect-questions-about-feminism](https://theguardian.com/technology/2019/sep/06/apple-rewrote-siri-to-deflect-questions-about-feminism)

61 [genderlessvoice.com](https://genderlessvoice.com)

n'est pas la solution unique à tous les problèmes de sexisme liées aux nouvelles technologies vocales. La neutralité ne permet pas nécessairement de créer des dispositifs non-sexistes. Le problème réside plutôt dans le fait d'assigner des voix genrées à des fonctions bien particulières et peu variées.

Écartons-nous de la question du féminisme et des enjeux de genre tout en maintenant le cap de cette réflexion sur la parole comme outil de représentation et de représentativité : et si on "donnait de la voix" à la Nature, pour mieux la défendre ?

Des réflexions mûrissent dans les murs des laboratoires de recherches ou des écoles, comme c'est le cas de l'organisation à but non lucratif Feminist Internet<sup>62</sup>, qui a pour objectif de "faire de l'internet un espace plus égal pour les femmes et les autres groupes marginalisés grâce à des pratiques créatives et critiques".

En 2018, un workshop "Designing a Feminist Alexa"<sup>63</sup> a été organisé avec 40 étudiant·e·s en design et en art de l'University of the Arts de Londres puis autour du chatbot F'xa<sup>64</sup> en 2019, pour fournir un guide

ludique sur les biais des intelligences artificielles.

Au cours du workshop "Designing a Feminist Alexa", les participant·e·s se sont interrogé·e·s sur ce qui faisait l'essence d'un assistant vocal féministe : qu'est-ce qu'une conversation féministe ? Existe-il une réponse féministe à la question "Alexa, quel temps fait-il aujourd'hui" ? Comment la valeur féministe d'une conversation entre un humain et une IA peut-elle être mesurée par les concepteur·ice·s ? Avec la production de huit projets aux contextes d'usages différents, les étudiant·e·s ont pu expérimenter le constat primaire de ce workshop : Alexa n'est pas conçue pour susciter des conversations féministes.

62 [feministinternet.com/projects/](https://feministinternet.com/projects/)

63 [drive.google.com/file/d/1v1rIT8dIA9mu-hvd-XfCCCCUQCujRhMOO/view](https://drive.google.com/file/d/1v1rIT8dIA9mu-hvd-XfCCCCUQCujRhMOO/view)

64 [about.f-xa.co](https://about.f-xa.co)

## ET SI ON IMAGINAIT UN ASSISTANT VOCAL POUR L'ACTIVISME ÉCOLOGIQUE ? ET

### SI LE VOCAL ŒUVRAIT POUR LA “RÉPUBLIQUE DU VIVANT” ?

Des assistants vocaux pour sensibiliser aux enjeux environnementaux ? L'idée peut sembler paradoxale, du moins provocatrice, quand on sait qu'une technologie vocale, notamment lorsqu'elle embarque une forme “d'intelligence” est loin d'être sans impact sur l'environnement. Mais pouvoir donner de la voix à des objets, c'est aussi pouvoir donner la parole au vivant muet. Et si la nature prenait la parole pour assurer sa propre défense ? C'est ce que simule la technologie présentée dans la design fiction “Végétalk”, dans un but de sensibilisation et de protection du patrimoine environnemental.

### Découvrez la bande sonore de la design fiction

Dans la lignée du “Parlement des choses”<sup>65</sup> de Bruno Latour, ou dans les pas de l'écri-

65 [bruno-latour.fr/sites/default/files/P-50-PARLEMENT-re-publication.pdf](http://bruno-latour.fr/sites/default/files/P-50-PARLEMENT-re-publication.pdf)

vain et journaliste Aymeric Caron, antispéciste - idéologie qui consiste à placer les intérêts des animaux au même niveau que ceux des humains - cette design fiction propose un futur où l'on s'avance (peut-être) vers une écoute et une défense de la parole végétale. La “République du vivant” (formule d'Aymeric Caron) ainsi constituée serait une “biodémocratie” où les arbres du poumon vert d'Amazonie, pour commencer, se verraient garantir leur préservation et la défense de leurs intérêts - à noter que la justice de Colombie a déjà reconnu la forêt amazonienne comme sujet de droit en avril 2018.

Les participant·e·s à la session de débat n'ont pas manqué de pointer le paradoxe inhérent à la design fiction - la défense du vivant par un moyen “trop technologique”, lui portant préjudice du même coup. Pour autant, l'impact sur les imaginaires s'est révélé fructueux, notamment grâce au pouvoir d'immersion offert par le son ; et on peut le comprendre, entendre un arbre qui se défend (même si en réalité “on le fait” se dé-



fendre) est une expérience puissante. C'est bien là la tentation (l'intérêt ? le fantasme ?) de l'anthropomorphisme : rendre intelligible le vivant qui parle une langue différente. Que celui ou celle qui n'a pas déjà rêvé de pouvoir donner la parole à son chien ou son chat me jette la première croquette.

## Donner “la parole” à celles et ceux qui en sont privé·e·s

L’anthropocentrisme est fréquent lorsqu’on cherche à communiquer avec d’autres espèces que la nôtre - il n’y a qu’à voir la manière dont on dépeint les extraterrestres dans la culture populaire, ils revêtent souvent une forme pseudo-humaine ou sont animés d’intentions qu’on peut catégoriser, l’alliance ou la guerre. Ce choix dans la représentation va aussi de paire avec un questionnement éthique capital lorsqu’il s’agit de simuler l’humain, puisqu’il peut y avoir méprise et engagement émotionnel.

Mais peut-être y-a-t-il dans certains cas un intérêt à exploiter cette tentation de l’anthropomorphisme. Dans un contexte d’état d’urgence environnemental, “faire parler” la nature c’est jouer sur le ressort émotionnel de l’humain pour enclencher une action qui n’a que trop tardé. Au-delà de l’action immédiate, cette parole artificielle peut-elle aider à un revirement de valeurs, sur le plan idéologique, en plaçant le vivant plus haut sur la scène juridique et démocratique ?

Au-delà de la parole qui est un mode d’expression parmi d’autres, peut-on imaginer des modes d’expression, à mi-chemin entre



l’humain et la nature, pour déployer un langage commun ?

### De l’activisme à la pédagogie : la nature augmentée

Si le contexte d’usage présenté dans la design fiction est évidemment provocateur - encore qu’il n’est pas si loufoque

quand on sait qu’en 2013 les arbres de la forêt amazonienne ont été équipés de téléphones capables d’appeler les autorités en cas d’abattage illégal<sup>66</sup> - on peut penser d’autres contextes où les technologies vocales pourraient avoir leur rôle à jouer auprès du vivant. Pourquoi pas des arboretums en

66 <https://www.20minutes.fr/plane-te/1082003-20130117-portables-arbres-amazonie>



richis de technologies vocales, où les arbres se présenteraient de manière autonome aux visiteur·se·s, dans un but pédagogique ? Des aventures ludiques, comme une chasse au trésor forestière où la végétation joue le rôle de maître du jeu ? Des performances poétiques, alliant technologie et végétal ? Peut-être est-ce l'occasion d'imaginer des rôles qui sortent de l'assistance de l'humain pour explorer des interactions voire des collaborations plus atypiques ?

### Le non-humain équipé pour la surveillance

A l'opposé de cette vision positivement inspirante, équiper le non-humain de technologies à même de capter des faits et gestes, et paroles humaines, c'est (s') exposer à d'autant plus d'agents potentiels de la surveillance. Les participant·e·s ont relevé ce danger, de même que le risque de générer de la nuisance sonore. Souhaiterait-on par exemple que les arbres de la ville soient équipés pour détecter les piétons qui traversent au feu rouge et capables de les rappeler à l'ordre en élevant la voix, devenant ainsi des agents de l'ordre mais aussi de l'humiliation publique ? On peut arguer que les technolo-

gies vocales, en se signalant, ont au moins le mérite de se rendre visibles, par rapport à d'autres beaucoup plus discrètes, comme la reconnaissance faciale. Peut-être justement pourrait-on l'envisager dans des rôles de “témoin de la transparence” dans la *smart city*, avertissant par exemple les citoyen·ne·s des capteurs qui se trouvent sur son chemin.

**Retrouvez les principaux arguments et points de friction soulevés lors de l'échange en Annexe 7.**

## ENJEUX LINGUISTIQUES : POUR INTERAGIR, IL FAUT SE COMPRENDRE !

Aujourd'hui la question de l'accessibilité des systèmes vocaux pour des voix aux accents régionaux, pour des langues peu parlées n'est pas adressée par les principaux fournisseurs d'assistants vocaux. L'entraînement des machines (*machine learning*) pour la reconnaissance vocale est en effet réalisé à partir d'échantillons de voix relativement réduits en termes de diversité linguistique, ce qui entraîne des problèmes conséquents d'utilisation par les personnes n'ayant pas un accent considéré comme standard par les concepteur·ice·s, que cela soit des accents régionaux ou des accents étrangers. C'est aussi le cas avec des mots, noms propres, qui ne sont pas reconnus par les système ; Wired <sup>67</sup>rapporte de nombreuses erreurs de compréhension comme celle-ci : la commande "Text Zahir" se transforme en "Text Zara here".

Plusieurs initiatives et acteurs se saisissent

<sup>67</sup> [wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/](https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/)

de cette question d'accessibilité.

La BBC avec Beeb<sup>68</sup>, son assistant vocal (lancement prévu en 2020) souhaite, en particulier, s'atteler à la question de l'accessibilité des assistants vocaux pour tout type d'accent régional. Ce constat a été fait aussi par certains concepteur·rice·s d'assistants vocaux dédiées à certaines populations. François Millet, concepteur de SKIPIT<sup>69</sup>, assistant vocal pour les personnes âgées, l'a expérimenté auprès de ces publics. Ces expérimentations ont été, dit-il, particulièrement révélatrices des problèmes de conformité des assistants vocaux classiques : la communication est presque impossible lorsque la machine ne connaît ni certaines intonations ni certains mots issus de dialectes locaux. Il a par ailleurs constaté que les personnes âgées comprenaient mieux leur assistant vocal lorsque celui-ci parlait avec un accent québécois. Le Washington Post<sup>70</sup> s'est également intéressé à la question et révèle de grandes disparités dans la compréhension des accents par les assistants vocaux selon les différentes régions américaines.

<sup>68</sup> [bbc.com/news/technology-49481210](https://www.bbc.com/news/technology-49481210)

<sup>69</sup> [44-studio.fr](https://www.44-studio.fr)

<sup>70</sup> [washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/](https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/)

Mozilla a lancé le projet Common Voice<sup>71</sup> pour réduire la fracture de la parole numérisée, faire en sorte que les données représentent la diversité de personnes réelles et permettre d'améliorer la reconnaissance vocale pour tous et toutes. Toute personne peut contribuer à ce projet en enregistrant sa voix sur le site web du projet.

Au-delà de la capacité des assistants vocaux à reconnaître des voix différentes parlant la même langue, se pose également la question du multilinguisme. Par exemple, Il existe des situations où des personnes peuvent échanger, interagir dans plusieurs langues à la fois. C'est le cas par exemple de familles qui parlent plusieurs langues à la maison ou qui accueillent des jeunes filles au pair, ou encore dans les hôtels et de façon générale, dans des pays ou provinces où cohabitent plusieurs langues (Québec, Belgique, Suisse ...). La plupart des assistants vocaux sont aujourd'hui programmés sur une seule langue à la fois, qu'il faut régler lors de l'initialisation de l'objet. Il est ensuite laborieux d'en changer, l'appareil n'étant pas prévu pour les usages cités précédemment. Il est même complexe de consulter des contenus qui ne

<sup>71</sup> [voice.mozilla.org](https://voice.mozilla.org)

sont pas issus de la langue programmée de l'assistant vocal, ce qui réduit considérablement toute forme d'ouverture culturelle.

Julia Velkovska et Moustafa Zouinar, respectivement sociologue et ergonomiste au Laboratoire SENSE (Sociology and Economics of Networks and Services) d'Orange Labs ont montré la complexité de l'interaction dite naturelle avec son assistant vocal. Ils ont étudié plusieurs situations dans lesquelles les assistants vocaux ne comprennent pas la question, la commande. Pour eux, il y a un travail central de l'utilisateur·rice pour arriver à se faire comprendre par la machine.

## QUELS SONT LES AUTRES ÉLÉMENTS QUI POURRAIENT FACILITER LES INTERACTIONS HUMAIN-MACHINE ?

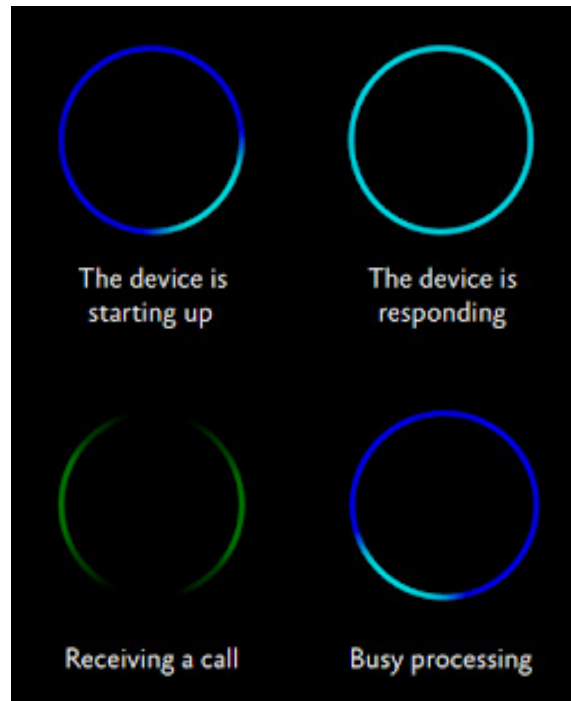


Image : des lumières d'Alexa et de leurs différentes significations.

D'autres éléments peuvent venir renforcer cette interaction, comme le rôle de la lumière, y compris pendant les silences. La conception d'un assistant vocal requiert l'intervention de plusieurs compétences en design : design graphique, design produit, design sonore, design de service etc... Comme le souligne Julien Drochon, designer, enseignant à l'ESA Pyrénées, la conception des assistants vocaux ne devrait pas se limiter à une prouesse technique mais devrait bel et bien servir l'usage. Les multiples interfaces (lumineuses, vocales, objet, graphique ...) des assistants vocaux doivent être complémentaires et se relayer dans différentes situations et avec divers utilisateur·trice·s. François Millet, concepteur de l'assistant vocal SKIPIT, souligne la nécessité de conserver une interface graphique qui puisse pallier à des déficiences de la voix humaine (maladie, mauvaises conditions sonores ...). Il propose également d'intégrer trois boutons principaux : "J'aime", "Je n'aime pas" et "Play" et 4 boutons de menu et volume. François Millet a fait le choix de laisser visible des "prises" pour l'utilisateur·e. Ces boutons sont, explique-t-il, des repères connus, notamment pour les personnes âgées, qu'elles peuvent associer aux boutons de la télécommande de leur télévision. Néanmoins, ce public n'est pas habitué à faire ses propres choix de pro-

grammes, mais à “zapper” parmi un choix réduit de programmes. La présence de boutons ne suffit donc pas toujours à rendre les objets lisibles pour certains publics.

Force est de constater que les enceintes connectées (Google Home, Amazon Echo, Djingo, Homepod ...) ont une forme similaire : un objet cylindrique aux couleurs sobres, qui a l'apparence d'un objet décoratif. Des contextes d'utilisation et d'usagers, amèneraient-ils à des formes différentes ? Questionner la forme de ces objets pourrait-elle être une entrée pour décroisonner les usages actuels et leur accessibilité ?



Lors de l'atelier de prototypage à l'Ensci, les étudiant·e·s ont interrogé la forme, l'identité de ces objets. En prenant l'exemple du gramophone, qui par sa forme révèle sa fonction et le dispositif associé, Selana Yao et Daniel Cadot ont proposé une approche didactique de l'assistant vocal. La requête vocale est décomposée en plusieurs éléments - de l'énoncé à la réponse - chacun ayant une fonction et une forme spécifiques :

1/ l'enregistreur de la voix qui émet un son capté par la machine (éven-

tail captant les sons, à l'image de l'oreille qui écoute) ;

2/ la machine intègre et émet son propre son permettant à l'utilisateur de comprendre que la machine digère la requête ;

3/ une fois la digestion faite par la machine, ces informations sont envoyées dans le cloud (tout ou partie) ;

4/ le résultat est restitué à l'utilisateur.

[recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/ok-oracular](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/ok-oracular)

Autour de cette question des formes de ces objets, il serait également possible de mieux comprendre le fonctionnement de cet objet comme par exemple révéler la présence d'un micro, savoir quand il est en mode écoute...

Ces nouveaux objets techniques, principalement opérés par les GAFAM, pourraient-ils nous en dire un peu plus sur leur fonctionnement ?

L'un des freins majeurs à l'adoption de ces enceintes connectées est la question de la confiance dans ces appareils vis-à-vis de leur vie privée. Selon l'étude Hadopi et CSA, 63% des utilisateur·rice·s réguliers d'enceintes connectées et des non utilisateur·rice·s estiment qu'elles peuvent constituer une menace pour leur vie privée. Comment ces questions de confiance et d'enjeu de vie privée peuvent-elles être prises en compte par les concepteur·rice·s?

**Territoire**

**5**

**Architecture**



## INVISIBILITÉ ET CONCEPTION

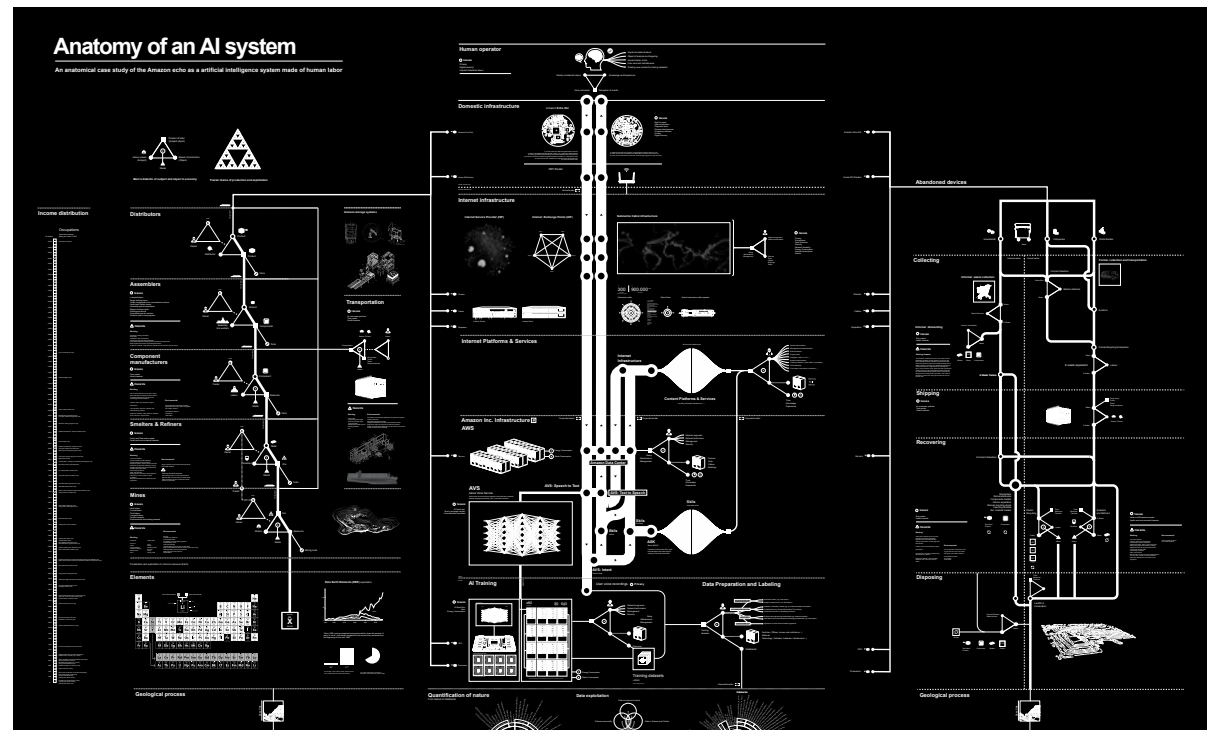
“Convenience is the most underestimated and least understood force in the world today.”<sup>72</sup>

La formulation d’une requête telle que “quel temps fait-il ?” implique, pour que le système vocal donne une réponse, des ressources naturelles, des lignes d’assemblage, des données et du travail humain pour entraîner la machine. Le système technique des assistants vocaux est complexe, comme l’ont montré Kate Crawford et Vladan Joler<sup>73</sup>. Leurs travaux permettent de rendre visible ce qui est caché par l’interface physique.

Les impacts environnementaux des assistants vocaux ainsi mis en lumière par ces chercheur·euse·s interrogent de fait la conception au regard de ces enjeux. Quelle est par exemple la consommation d’énergie supplémentaire nécessaire pour toute commande vocale domotique qui se substituerait à une commande manuelle simple ? Un

<sup>72</sup> [nytimes.com/2018/02/16/opinion/sunday/tyranny-convenience.html?mtrref=t.co&assetType=opinion](https://www.nytimes.com/2018/02/16/opinion/sunday/tyranny-convenience.html?mtrref=t.co&assetType=opinion)

<sup>73</sup> [anatomyof.ai](https://anatomyof.ai)



[anatomyof.ai](https://anatomyof.ai)

exemple emblématique est celui de l’activation d’un interrupteur par la voix. L’installation d’un hub domotique avec commande vocale devrait prendre en considération cet enjeu écologique, avec un calcul du coût et/ou du gain énergétique potentiel. Il nous semble important de poser ces questions dès la conception, avant un déploiement généralisé de ces dispositifs dans l’espace pri-

vé ou public. Récemment, des chercheurs<sup>74</sup> de l’Université du Massachusetts ont calculé l’impact carbone lié à l’entraînement d’un modèle de *machine learning* pour le traitement du langage naturel. Ils ont conclu à une équivalence de 125 allers-retours New York/Pékin en avion.

<sup>74</sup> [arxiv.org/pdf/1906.02243.pdf](https://arxiv.org/pdf/1906.02243.pdf)

On s'aperçoit que la forme de ces objets, qui semblent si anodins, masque une complexité technique invisible et un coût social, environnemental et économique significatif. Aux enjeux environnementaux s'ajoutent les enjeux sociaux. Nous sommes encore bien loin de la machine auto-apprenante, intelligente, que les assistants vocaux nous promettent, à l'image de ce témoignage <sup>75</sup> qui explique comment nos conversations avec les machines sont écoutées, retranscrites, analysées par des humains, véritables "dresseurs d'IA" invisibles. Les technologies nécessaires au bon fonctionnement des interfaces vocales ont un coût humain. Les interactions vocales avec une machine sont écoutées, retranscrites, corrigées par des humains. Quelle est la part de travail humain dans ces technologies ? Comment qualifier ces nouveaux emplois ? Pour une intelligence artificielle "faible", combien de travailleur·euse·s invisibles et précaires ? Le service "Google Duplex", déployé aux Etats-Unis, appelle au travail humain. Ce service propose de déléguer à l'assistant vocal Google les prises de rendez-vous ou d'établir des réservations pour le compte de l'utilisateur·rice. Or, il ne fonctionne pas à 100%

<sup>75</sup> [laquadrature.net/2018/05/18/temoin\\_cortana/](http://laquadrature.net/2018/05/18/temoin_cortana/)

en autonomie mais grâce à un pourcentage non négligeable d'employé·e·s qui passent les appels à la place des machines. Dans le même registre, Amazon dispose d'une plateforme Alexa Answers, où des utilisateur·rice·s d'Alexa peuvent contribuer - sans être rémunéré·e·s - à l'amélioration de l'assistant vocal en répondant aux questions d'autres utilisateur·rice·s... Encore une fois, nous interrogeons le sens de ces nouveaux services, devons-nous accepter de déléguer des micro-tâches à des machines qui finalement les "délèguent" à des humains ? Pouvons-nous nous satisfaire d'un service "humain-machine-humain" ? Peut-on encore parler dans ce cadre d'innovation de service ou d'innovation technologique ?

### Assistant vocal et nouvelles invisibilités domestiques

La fonction d'assistant·e est couramment attribuée à un humain qui "aide quelqu'un dans l'exercice de ses fonctions en se tenant auprès de lui"<sup>76</sup>. Cette fonction s'inscrit dans la continuité d'un ensemble de fonctions et d'emplois généralement précaires et néces-

<sup>76</sup> Définition du CNRTL

sitant un niveau de qualification bas. Quand les domestiques travaillent dans le cadre privé du domicile, les assistant·e·s agissent plutôt dans le contexte de l'entreprise. Le terme d'assistant·e est devenu un terme générique pour définir une variété d'emplois et il a conquis l'espace domestique au travers des dispositifs comme Google Home ou Amazon Echo, des voix artificielles intégrées dans différentes pièces, pour répondre aux besoins des habitant·e·s du lieu. Ces dispositifs vocaux s'intègrent dans l'écosystème de la "maison intelligente", aux côtés de nombreux objets de domotique, allant des caméras de surveillance à la climatisation et au système d'éclairage connecté. L'implémentation d'une interface vocale sur ces dispositifs nourrit un imaginaire de la maison que Saul (Pia) Pandelakis, enseignant-chercheur à l'Université Toulouse - Jean Jaurès, décrit comme "un environnement d'objets pensés comme de petits esclaves qui, besogneux, s'affairent pour garantir le confort de la maison"<sup>77</sup>. Le terme "esclave" renvoie ici non pas à la condition machinique de ces objets mais plutôt à la reproduction d'une culture

<sup>77</sup> Saul (Pia) Pandelakis, "Done by app" : du design de services au quadrillage du réel. MEI - Médiation et information, L'Harmattan, 2017, Design et Communication. repéré : [mei-info.com/revue/40/63/done-by-app-du-design-de-services-au-quadrillage-du-reel/](http://mei-info.com/revue/40/63/done-by-app-du-design-de-services-au-quadrillage-du-reel/)



de la domestication que l'on s'emploie à dissimuler, auparavant grâce à des systèmes architecturaux - comme les escaliers de service, les chambres de bonne, les tapis pour masquer le bruit des pas<sup>78</sup>- et aujourd'hui par des écosystèmes numériques complexes où l'humain conserve une place indispensable. Saul (Pia) Pandelakis décrit plus amplement dans l'article "Done by app"<sup>79</sup> la manière dont les dispositifs numériques de conciergerie, de la livraison à domicile aux services à la personne, tendent à invisibiliser un travail humain de plus en plus précaire. Il est donc nécessaire de questionner le terme "d'assistant vocal", décrivant aujourd'hui des dispositifs comme Siri, Alexa, Cortana ou Google Home, tant parce qu'il clôture les imaginaires autour de cette technologie vocale que parce qu'il reproduit des schémas de domination au travers de dispositifs numériques.

Ces enjeux interrogent notre capacité en tant que citoyen·ne à pouvoir choisir à bon escient, d'utiliser ou non ces technologies, en ayant la connaissance des enjeux sociaux et environnementaux identifiés précé-

78 Ibid.

79 Ibid

demment. Quelle lisibilité nous est offerte ? Quelles clefs de lecture pour une conception responsable ? La délégation de micro-tâches à une machine présente un coût social et environnemental, souhaitons-nous que ce modèle devienne dominant et pour quels bénéfices ? Devons-nous résister à la "tyrannie de la commodité" comme nous y incite le juriste américain Tim Wu<sup>80</sup> ou l'artiste américaine Lauren McCarthy ?

80 [nytimes.com/2018/02/16/opinion/sunday/tyranny-convenience.html?mtrref=t.co&assetType=opinion](https://www.nytimes.com/2018/02/16/opinion/sunday/tyranny-convenience.html?mtrref=t.co&assetType=opinion)

# LAUREN



Lauren McCarthy propose dans sa performance "LAUREN"<sup>81</sup> de devenir une version humaine d'Alexa d'Amazon, à distance, pendant plusieurs jours, elle surveille le domicile d'une personne et se présente comme une IA améliorée. L'objet de cette performance vise à mettre en lumière les tensions entre automatisation et travail humain, entre intimité et surveillance, entre commodité et perte de contrôle.

81 [lauren-mccarthy.com/LAUREN](https://lauren-mccarthy.com/LAUREN)

## Invisibilité et privacy

En France, “les utilisateur·rice·s sont [...] pour deux tiers environ à déclarer craindre pour la confidentialité de leurs données personnelles (62 %) et à estimer que les enceintes connectées constituent une menace pour leur vie privée (61 %).<sup>82</sup>” Les chiffres<sup>83</sup> avancés aux Etats-Unis sont nettement moins élevés, seuls 26% des propriétaires d’enceintes connectées se sentent “très concernés par les enjeux autour de la vie privée”.

Les principaux assistants vocaux connus du grand public (Google Home, Alexa, Cortana, Siri...) s’appuient sur des systèmes techniques de reconnaissance vocale, qui impliquent une écoute passive des utilisateur·rice·s et nécessitent généralement un transfert de données audio vers les serveurs de ces grands acteurs. Il est techniquement admis, du fait de la conception de ces assistants vocaux et de l’écoute passive, que des données vocales, des conversations peuvent se retrouver par erreur sur les ser-

82 [hadopi.fr/sites/default/files/sites/default/files/ckeditor\\_files/2019\\_05\\_24\\_Assistants\\_vocaux\\_et\\_enceintes\\_connectees\\_FINAL.pdf](https://hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2019_05_24_Assistants_vocaux_et_enceintes_connectees_FINAL.pdf)

83 [voicebot.ai/wp-content/uploads/2019/03/smart\\_speaker\\_consumer\\_adoption\\_report\\_2019.pdf](https://voicebot.ai/wp-content/uploads/2019/03/smart_speaker_consumer_adoption_report_2019.pdf)

veurs des opérateurs. La CNIL<sup>84</sup> recommande à juste titre d’éteindre les enceintes connectées lorsque l’utilisateur·rice n’en a pas l’usage. Depuis que nous avons débuté le travail sur ce projet, les témoignages se sont multipliés sur les pratiques d’écoute de Google, Amazon, Apple et Microsoft, tous reconnaissant écouter et enregistrer régulièrement les conversations des utilisateur·rice·s, sous couvert d’amélioration des systèmes vocaux. Et ce, pour entraîner les systèmes à la reconnaissance vocale, pour corriger et améliorer la pertinence de la réponse. L’écoute serait une pratique largement partagée par ces concepteur·rice·s. Quelle confiance accorder à ces dispositifs ? Sommes-nous condamné·e·s à sacrifier une part de nos vies privés contre l’usage de ces dispositifs vocaux ? Quels sont les modèles alternatifs de conception, *privacy by design* ? Devons-nous détourner, contourner ces dispositifs ? Quelles stratégies leur opposer ?

84 [cnil.fr/fr/enceintes-intelligentes-des-assistants-vocaux-connectes-votre-vie-privée](https://cnil.fr/fr/enceintes-intelligentes-des-assistants-vocaux-connectes-votre-vie-privée)

## ARCHITECTURES ALTERNATIVES & DIDACTIQUES

### Des modèles alternatifs existent<sup>85</sup>

Snips<sup>86</sup>, start-up française spécialiste des logiciels de reconnaissance vocale embarquée dans les objets connectés, propose une solution de conception alternative, plus respectueuse de la vie privée des utilisateur·rice·s, avec un traitement des données vocales en local, pour commander des objets par la voix. Les requêtes sont organisées par domaines d’usages contextualisés. Tout est traité en local, à l’exception des requêtes hors du domaine d’usage prévu. Le modèle est entraîné à partir de données achetées ou *crowdsourcées*, permettant de reconnaître et comprendre une requête dans le domaine d’usage. Par exemple, le domaine d’usage autour d’un appareil domotique, le domaine

85 La start up Snips a été rachetée en novembre 2019 par Sonos, fabricant d’équipement audio, au moment de la rédaction de ce livrable Hypervoix. Ce rachat présage-t-il des modifications sur cette question de la protection privée : abandon ou renforcement de la vie privée ?

86 [internetactu.net/2019/05/24/hypervoix-12-concevoir-les-interactions-vocales/](https://internetactu.net/2019/05/24/hypervoix-12-concevoir-les-interactions-vocales/)

d'usage autour de la musique, de la météo... Snips limite les requêtes à des domaines d'usages contextualisés afin de traiter les données en local.

Mozilla avec son initiative Common Voice<sup>87</sup> veut ouvrir l'accès aux jeux de données vocales pour casser les barrières d'accès à la voix pour les développeur·se·s, la dépendance aux services tiers et pour les utilisateur·rice·s, la dépendance aux biais du marché, aux langues dominantes, à la non prise en compte des accents. Il faut des milliers d'heures d'enregistrement pour alimenter un modèle de reconnaissance vocale. La puissance de calcul ainsi que la quantité de données nécessaires aux modèles de reconnaissance vocale limitent l'accès à cette technologie, les acteurs se trouvent liés à des services tiers dominants. L'objectif de contribution par langue est fixé à 10 000 heures, soit plusieurs mois de voix en continu. Pour les concepteur·rice·s, un accès ouvert de gros jeux de données vocales. Pour les utilisateur·rice·s, la possibilité de contribuer dans sa langue, avec son accent régional, en fonction de son âge. Common Voice ouvre les jeux de données vocales, casse les biais d'accès et prend en compte la diversité des publics.

Snips nous rappelle que le modèle "cloud", pour des commandes vocales dans des usages contextualisés, est contournable. Common Voice propose d'ouvrir des jeux de données vocales *crowdsourcées*.

### D'autres initiatives

L'Inria, coordinateur du programme de recherche européen COMPRISE<sup>88</sup> initié en 2019, vise à concevoir une nouvelle génération d'outils d'interaction vocale *privacy by design*, plus inclusifs et propose ainsi une alternative dans la conception des interfaces vocales.

Le Voice Lab<sup>89</sup>, initiative française pour fédérer l'écosystème européen autour de la voix comme interface, peut être la chambre d'écho des modèles alternatifs. L'objectif premier affiché par les acteurs du VoiceLab

vise à "constituer et diffuser des corpus vocaux pour entraîner des algorithmes IA pour la reconnaissance de la parole", pour se libérer de la dépendance des systèmes techniques vocaux propres aux GAFAM/BATX.

La mission Etalab<sup>90</sup> a initié fin octobre 2019 une démarche contributive s'inspirant de Common Voice, pour constituer un jeu de données francophones sur la base de questions-réponses pour l'entraînement des IA francophones afin "d'améliorer la performance d'agents conversationnels"<sup>91</sup> notamment.

88 [inria.fr/actualite/mediacenter/vers-la-prochaine-generation-d-outils-d-interaction-vocale](https://inria.fr/actualite/mediacenter/vers-la-prochaine-generation-d-outils-d-interaction-vocale)

89 [levoicelab.org](https://levoicelab.org)

90 [etalab.gouv.fr/qui-sommes-nous](https://etalab.gouv.fr/qui-sommes-nous)

91 [etalab.gouv.fr/ia-decouvrez-et-participez-au-projet-piaf-pour-des-ia-francophones](https://etalab.gouv.fr/ia-decouvrez-et-participez-au-projet-piaf-pour-des-ia-francophones)

87 [vimeo.com/304803157](https://vimeo.com/304803157)

## Pratiques (ludiques) de détournement des interfaces

Des designers s'emparent des enjeux de *privacy*, à l'image du projet Alias qui vise à redonner aux utilisateur·rice·s un moyen de contrôler ces dispositifs d'écoute intrusive.



Projet Alias

Le projet Alias<sup>92</sup>, développé par Borek Karman et Topp Tore Knudsen, est un “parasite”, un couvercle imprimé en 3D qui se fixe au sommet d’une enceinte connectée; à l’intérieur, un microphone et deux haut-parleurs produisent un bruit blanc qui empêche le haut-parleur cible de s’activer. L’utilisateur·rice paramètre lui-même le mot clé

92 [bjoernkarman.dk/project\\_alias](http://bjoernkarman.dk/project_alias)

d’activation pour reprendre la main sur le dispositif d’écoute et son déclenchement. Erlend Prendergast, étudiant en design de l’Ecole d’Art de Glasgow, propose une approche alternative et satirique de protection de nos vie privées numériques, avec son projet CounterBug.<sup>93</sup> Il décrit CounterBug comme “une famille d’accessoires spéculatifs destinés à confondre les algorithmes d’Alexa. Chaque accessoire répond à une forme différente de paranoïa de surveillance : un pour censurer la langue de l’utilisateur·rice afin de ne pas avoir de problèmes avec la NSA, un autre pour perturber les algorithmes publicitaires d’Amazon, un autre encore qui discute avec Alexa de sujets vertueux alors que l’utilisateur·rice est chez lui.”

Ces deux projets, principalement à visée pédagogique voire ludique, alertent les utilisateur·rice·s sur les enjeux de vie privée en cas d’utilisation d’enceintes connectées.

David Gaborieau<sup>94</sup>, sociologue, dans son livre “Le Monde en pièces”, paru aux éditions La lenteur, consacre tout un chapitre

93 [erlendprendergast.com/CounterBug](http://erlendprendergast.com/CounterBug)

94 «Le logiciel qui murmure à l’oreille des pros», in *Le Monde en pièces, pour une critique de la gestion*, vol. 2, «Informatiser», édition La lenteur, David GABORIEAU, 2019

sur le “voice picking”, technologie développée à partir des années 1990 chez Ford Motors puis chez WalMart, qui est un système de guidage vocal utilisé notamment dans les entreprises de logistique par les préparateur·trice·s. Il décrit comment les préparateur·trice·s se sont emparé·e·s du système de guidage pour mieux le détourner. En mémorisant les codes détrompeurs permettant d’identifier les produits dans un entrepôt, les préparateur·trice·s “hackent le système” et composent eux·elles-mêmes leur palette pour reprendre la main. Ils·elles jouent sur le système de reconnaissance vocale, en transformant les codes d’emplacement par des mots détournés : “4-1 devient catin, 5-1 devient sapin, 2-2 devient neuneu”, une forme ludique pour tromper la machine et détourner la rationalisation du métier.

Face à la multiplication des dispositifs techniques de reconnaissance, de détection de présence, d’écoute, d’enregistrement, une part de nos identités numériques nous échappent. Les données biométriques s’ajoutent désormais à la liste des données personnelles constituant nos identités numériques et ayant vocation à nous authentifier et à nous identifier dans nos interactions avec ces dispositifs.

Inventerons-nous de nouvelles identités numériques pour préserver nos vies privées, inventerons-nous de nouveaux langages ? Vers de nouvelles identités numériques, sources d'invisibilité numérique temporaire ou d'identification «atténuée», un pseudonymat de nos voix ?

« L'utilisateur est mis en visibilité, alors que le système, lui, masque son fonctionnement ».<sup>95</sup> Même si des guides apparaissent pour apprendre à paramétrer les assistants vocaux et limiter les intrusions dans nos vies privées, de quelle marge d'autonomie et de compréhension dispose un individu confronté directement ou indirectement à ces dispositifs ? Ces projets présentés ci-dessus sont des prototypes à but didactique non commercialisés. L'individu reste pour l'instant seul, face à des systèmes sur lesquels il a peu de contrôle, peu de compréhension, peu de visibilité. Or la relation de confiance entre l'utilisateur·rice et l'objet commandé par la voix est essentielle à l'adoption de ces dispositifs. La CNIL rappelle quelques principes<sup>96</sup> de vigilance sur la protection de la vie pri-

vée dans l'usage des enceintes connectées ou des assistants vocaux.

Comment rendre visible la voix ou l'écoute ? Comment savoir si un objet est équipé d'un micro ou doté d'une voix ? A mesure que les systèmes vocaux se répandront dans l'espace privé, public, en mobilité, opposera-t-on un droit d'information, de signalement en cas de présence d'un dispositif d'enregistrement des voix, des sons ? Le développement de nouvelles interfaces numériques dans l'espace privé ou public devrait s'accompagner par la mise en place de nouvelles signalétiques, régulées et discutées. Si l'individu ne peut échapper à ces dispositifs, il devrait a minima avoir connaissance de leur présence. Ces interfaces devraient a minima "se signaler" afin d'éviter des écoutes non souhaitées, des interruptions inopportunes et une surveillance généralisée.

---

<sup>95</sup> [internetactu.net/2018/10/18/privacy-by-design-concevoir-pour-preserver-la-vie-privee/](https://internetactu.net/2018/10/18/privacy-by-design-concevoir-pour-preserver-la-vie-privee/)

<sup>96</sup> [service-public.fr/particuliers/actualites/A13218](https://service-public.fr/particuliers/actualites/A13218)



***En synthèse***

Nous souhaitons en synthèse partager des points de vigilance, des enseignements à partir de ce travail exploratoire et proposer des recommandations pour les concepteur·rice·s mais aussi les utilisateur·rice·s. de ces interfaces vocales.

## Enjeux d'intégration

- » Avec l'accroissement des objets connectés et des interactions vocales, anticiper le risque de cacophonie sonore et verbale dans tout lieu partagé.

### TERRITOIRE 1

- » Travailler sur la notion de hub de régulation -ou tableau de bord- afin de faciliter la gestion des objets connectés parlants, leur synchronisation, et de minimiser les conflits d'usage (e.g. plusieurs dispositifs qui répondent à une même commande utilisateur).

### TERRITOIRE 1

- » Travailler sur l'usage collectif des assistants vocaux et les priorités d'accès dans un lieu partagé.

### TERRITOIRE 1

## Enjeux de transparence

- » Rendre visible et lisible les systèmes vocaux : tant par des approches didactiques que par un travail sur l'appropriation possible de ces objets techniques (e.g. tableau de bord).

### TERRITOIRES 4 & 5

- » Citer les sources ou à minima rendre possible la consultation des sources à l'origine de l'information.

### TERRITOIRES 1 & 2

- » Permettre aux utilisateur·rice·s de toujours garder à l'esprit qu'ils interagissent avec une machine grâce à des éléments signalétiques et/ou des choix de conception spécifiques concernant la figure de l'assistant vocal.

### TERRITOIRES 3 & 4

- » Travailler sur des formes de paramétrage permettant aux utilisateur·rice·s de garder le contrôle sur leur relation aux as-

sistants vocaux et d'en avoir un usage responsable et consenti (lisibilité, compréhension des enjeux, prise en main).

### TERRITOIRES 3 & 5

## Enjeux de protection de la vie privée

- » Concevoir des dispositifs vocaux respectueux de la vie privée des individus. Cela peut passer par le développement de solutions private by design : des dispositifs non connectés, des systèmes d'anonymisation des voix par encodage, mais aussi par un travail approfondi sur le consentement.

### TERRITOIRES 1 & 5

- » Envisager, en fonction des usages autour des données vocales, une évolution du droit : quel consentement éclairé de l'individu sur ses données vocales ?

### TERRITOIRE 1

## Enjeux attentionnel et cognitif

- » Travailler sur l'espace attentionnel de l'utilisateur·rice : ménager sa capacité d'écoute et de concentration, éviter les messages longs et inappropriés, privilégier des tours de parole courts et laisser l'utilisateur·rice décider du bon moment pour interagir en évitant de faire parler la machine de façon proactive.

### TERRITOIRE 3

## Enjeux d'inclusion, de représentativité et de protection des publics fragiles

- » Prendre en compte l'accessibilité dans la conception de dispositifs vocaux : prendre en compte les différents niveaux de langage des utilisateur·rice·s, les accents, le multilinguisme et les troubles de l'élocution (bégaiement etc).

### TERRITOIRE 4

- » Prendre en compte l'âge et la maturité des utilisateur·rice·s dans la conception des dispositifs vocaux, dont la capacité à garder de la distance face à ce dispositif très humanisé.

### TERRITOIRES 2 & 4

- » Prendre en compte la courbe d'apprentissage des utilisateur·rice·s : le niveau d'autonomie et le temps d'utilisation.

### TERRITOIRES 2 & 4

## Enjeux de diversité

- » Garantir un accès ouvert à des corpus vocaux pour l'entraînement des modèles de traitement du langage naturel afin de permettre le développement d'un écosystème d'acteurs.

### TERRITOIRES 4 & 5


- » Garantir la diversité des sources de données vocales associées aux enceintes connectées comme le recommande l'Arcep.

### TERRITOIRE 2

- » Former les designer·euse·s aux enjeux de conception de dispositifs vocaux (*machine learning*, rôle des données, etc.).

### TERRITOIRES 2 & 3





***Les angles  
morts, champs  
de recherche***

Dans ce travail exploratoire, certaines dimensions de l'interaction mais aussi de l'impact de ces nouvelles interfaces n'ont pas été développées. Nous souhaitons tout de même les mentionner et permettre ainsi à d'autres de s'en saisir. Il en est de même du côté de la recherche.

- » Quelle alternative au système à réponse unique de type *Answer Engine* (moteur de réponse) et quelle(s) proposition(s) en faveur d'une forme d'interaction davantage au service de l'apprentissage plutôt que du gain de temps immédiat ?

#### **DESIGN FICTION "LE BVP", TERRITOIRE 2**

- » À l'échelle individuelle, l'assistant vocal est proposé comme un agent facilitateur permettant, entre autre, de mieux organiser son temps et ses actions. Que pourrait devenir son rôle au sein d'un collectif ? Comment en faire une éventuelle aide à la coordination et la coopération, sans risquer une perte d'autonomie du collectif ? Comment faciliter sans désintermédiaire ?

#### **DESIGN FICTIONS "LEADERBOT" ET "ARLETTE", TERRITOIRE 1**

- » Les assistants vocaux ainsi que les terminaux permettant d'interagir avec

eux (smartphone, smartspeaker, etc) sont connus pour leur faible résistance au hacking, que ce soit par le déclenchement à distance d'action ou le décryptage d'enregistrements utilisateur·rice·s. Quelles solutions de protection apporter aux utilisateur·rice·s, aux personnes tiers (également susceptibles d'être écoutées) mais aussi aux entreprises et aux institutions ? Qu'est-ce qui relève de choix collectifs et qu'est-ce qui relève du consentement individuel ?

- » L'apparition des assistants vocaux ne saurait effacer le versant visuel et tactile de nos habitudes d'interactions, ni répondre à tous nos besoins (lire, regarder une image, etc). L'avenir semble être à la superposition des dispositifs mais aussi à la multimodalité des expériences. Quel place pour le vocal ? Quelle complémentarité par rapport aux autres modes cognitif et sensoriel ? Quelle versatilité des

contenus (un contenu peut être proposé en audio, en vidéo, transcrit sous forme de texte, en format long ou court, etc) ?

#### **TERRITOIRE 3**

- » Quelle mesure des impacts sociaux et écologiques liés à la conception et l'utilisation des dispositifs vocaux ?

#### **TERRITOIRE 5**

- » Dans quelle mesure les assistants vocaux peuvent-ils jouer le rôle de porte-voix ou porte-parole, au service de revendications individuelles ou collectives ? Quelles gouvernance et organisation de l'information collectée (pas de neutralité possible) ?

#### **DESIGN FICTIONS "ARLETTE", TERRITOIRE 1**

- » Doit-on pouvoir ajuster le niveau de relation que nous souhaitons avoir avec un assistant vocal (rapport amical, fonctionnel, etc) en fonction de ses envies et ses besoins ? Que serait un "bon

usage” de la relation conversationnelle ?  
Et en fonction de quel usage un rapport d’autorité (figure de l’enseignant) ou un rapport de soumission (figure de major-dome) peuvent se justifier ?

**TERRITOIRES 3 & 4**

- » Dans quel mesure les technologies vocales peuvent permettre de “donner la parole” ou “donner une voix” à des formes d’existence non humains tel que les animaux ou les végétaux, que ce soit dans un but de survie ou de divertissement ? Comment encadrer ce type de démarche ?

**DESIGN FICTIONS “VÉGÉTALK”,  
TERRITOIRE 4.**

A large, stylized letter 'A' graphic in a light teal color, positioned on the left side of the page. It has a thick, blocky appearance with a slight shadow effect.

# ***Annexes***

## ANNEXE 1 : UN RÉSUMÉ HISTORIQUE DES ASSISTANTS VOCAUX

Les technologies vocales ont déjà une histoire riche avec de nombreux dispositifs : le téléphone bien sûr, développé durant la seconde moitié du XIXe siècle par Philippe Reis, Antonio Meucci puis Alexander Graham Bell, permet de communiquer à distance de façon vocale ; le haut parleur breveté le 10 décembre 1877 par Werner von Siemens, permet de diffuser la voix ; le dictaphone mise au point en 1907 par la Columbia Graphophone Company fait suite au phonographe développé dès le milieu du XIXe siècle, permet d'enregistrer la voix ; et enfin, le développement des voix de synthèse permettant d'imiter la voix, dès 1791 avec la *Speaking Machine* de Wolfgang von Kempelen, mais surtout à partir du XXe siècle avec les travaux de recherche des laboratoires Bell puis de l'IRCAM à partir de 1977. Les assistants vocaux opèrent une synthèse de ces différentes technologies : ils écoutent et enregistrent des requêtes vocales et diffusent des réponses via des voix de synthèse de plus en plus réalistes.

“Le premier assistant vocal, Audrey, est né aux Bell Laboratories en 1952 : il était capable de reconnaître les chiffres de 0 à 9 afin de faciliter la commutation téléphonique. Il a été suivi de nombres d'amélioration, comme la Shoebox d'IBM (1961), du premier programme de chatbot, Eliza, de Joseph Weizenbaum en 1964 qui visait à simuler sous forme textuelle, un psychothérapeute. Puis de programmes de la Darpa et d'une machine à écrire vocale d'IBM dans les années 80 qui donneront naissance aux premiers systèmes de reconnaissance vocale. En 1987, l'un des grands spécialistes du sujet, Frederick Jelinek explicite les enjeux de ces recherches en posant le principe d'une interface « naturelle », sans que cette caractéristique ne soit jamais pleinement explicitée.

Les véritables applications commerciales vont commencer à décoller avec le développement de l'informatique personnelle. En 1997, Dragon lance la première version de son logiciel de reconnaissance vocale, qui transcrit la voix en mots. Le boom actuel des outils de reconnaissance vocale repose sur les progrès réalisés depuis et qui s'incarnent au début des années 2010, avec Watson d'IBM (2011), Siri d'Apple (2011), Google Now (2012), Cortana de Microsoft (2013) des interfaces qui vont entrer en interaction

avec les fonctions du téléphone pour proposer des réponses à des questions vocales. Puis, ces assistants intégrés aux téléphones vont s'externaliser dans des objets dédiés pour une utilisation domestique, comme les enceintes connectées avec Alexa d'Amazon (2014), Google Home (2016), HomePod d'Apple (2018), Portal de Facebook (2018).”

*Extrait de l'article “Si les assistants vocaux sont la solution, quel est le problème ?”<sup>97</sup>*

Hubert Guillaud, InternetActu

---

<sup>97</sup> [internetactu.net/2018/12/19/vox-machines-12-si-les-assistants-vocaux-sont-la-solution-quel-est-le-probleme/](https://internetactu.net/2018/12/19/vox-machines-12-si-les-assistants-vocaux-sont-la-solution-quel-est-le-probleme/)

## ANNEXE 2 : VEILLE, ARTICLES TOUT AU LONG DU PROJET

### Le fil de veille collaboratif HyperVoix

» [groups.diigo.com/group/hyper\\_voix](https://groups.diigo.com/group/hyper_voix)

Avec plus de 400 articles référencés

### Les comptes-rendus des journées d'étude

La journée d'étude par le Master Design transdisciplinaire, cultures et territoires de l'université de Toulouse - décembre 2018

Programme :

» [recherche.esad-pyrenees.fr/vox-machines/journee-d-etude-1/programme-detaille](https://recherche.esad-pyrenees.fr/vox-machines/journee-d-etude-1/programme-detaille)

Articles compte rendu de cette journée :

- » [internetactu.net/2018/12/19/vox-machines-12-si-les-assistants-vocaux-sont-la-solution-quel-est-le-probleme/](https://internetactu.net/2018/12/19/vox-machines-12-si-les-assistants-vocaux-sont-la-solution-quel-est-le-probleme/)
- » [internetactu.net/2018/12/20/vox-machines-22-a-qui-les-interfaces-vocales-sadaptent-elles/](https://internetactu.net/2018/12/20/vox-machines-22-a-qui-les-interfaces-vocales-sadaptent-elles/)

La Journée d'étude autour des enjeux de conception des assistants vocaux - avril 2019 Square Paris

Programme et intervenants :

» [recherche.esad-pyrenees.fr/vox-machines/journee-d-etude-2](https://recherche.esad-pyrenees.fr/vox-machines/journee-d-etude-2)

Articles compte rendu de cette journée :

- » [internetactu.net/2019/05/24/hyper-voix-12-concevoir-les-interactions-vocales/](https://internetactu.net/2019/05/24/hyper-voix-12-concevoir-les-interactions-vocales/)
- » [internetactu.net/2019/05/27/hyper-voix-22-les-enjeux-de-la-voix-au-dela-de-la-voix/](https://internetactu.net/2019/05/27/hyper-voix-22-les-enjeux-de-la-voix-au-dela-de-la-voix/)

## ANNEXE 3 : CARTOGRAPHIE DES PROMESSES ET FREINS DES ASSISTANTS VOCAUX

Dans le cadre de notre exploration du sujet HyperVoix, nous avons proposé une cartographie des freins au regard des promesses énoncées par les concepteur-ric-e-s ou tout simplement forgées par les imaginaires.

### Promesse 1

#### “LES MAINS LIBÉRÉES”

- » #hyper-attentionnel
- » #capacité de mémorisation
- » #navigation - méconnaissance des fonctionnalités
- » #complexité cognitive

### Promesse 2

#### “L’ASSISTANT PERSONNEL / LE MAJORDOME / LE COMPAGNON”

- » #défiance
- » #gestion des identités
- » #diversité des publics
- » #interruption inopportune
- » #personnalité lisse, ennuyeuse
- » #isolement

### Promesse 5

#### “L’UBIQUITÉ” - MULTI-MODALITÉ

- » #concurrence avec le tactile
- » #cacophonie
- » #superposition d’informations

### Promesse 4

#### “L’INTERACTION NATURELLE”

- » #interaction initiée avec un mot d’activation
- » #formulation de la requête / déshumanisation de l’échange
- » #réponse unique
- » #concept de l’alignement / accommodation
- » #écarts culturels

### Promesse 3

#### “L’IA”

- » #non pertinence émotionnelle
- » #non pertinence de la réponse
- » #privacy / data
- » #consentement
- » #surveillance
- » #enfermement #biais
- » #invisibilité
- » #impacts sociaux & éco.
- » \*deep learning \*explicabilité
- » #régulation - neutralité du net

## ANNEXE 4 : WORKSHOP ENSCI

Workshop piloté par Julien Drochon et Anthony Masure, designers.

12 étudiants / 4 jours pour proposer des maquettes à partir de la journée d'étude / 5 projets.

### > Featuring

Comment offrir une part de créativité à la machine ? Interroger l'interaction :

[recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/featuring-computing](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/featuring-computing)

### > Le bon, la boîte et le truand

Unicité de la réponse des AV, jusqu'à un appauvrissement culturel et de la pensée ?

[recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/le-bon-la-boite-et-le-truand](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/le-bon-la-boite-et-le-truand)

### > Nosferatu recipe

La captation des données par les assistants vocaux, comment sensibiliser les individus aux enjeux de vie privée liée aux AV ?

[recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/charles](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/charles)

### > Memory Print

Archivage des contenus, mémoire, traces, comment matérialiser la voix ?

[recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/memory-print](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/memory-print)

### > L'Oracle

Une approche didactique de la voix

[recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/ok-oracular](https://recherche.esad-pyrenees.fr/vox-machines/workshop/projets-etudiants/ok-oracular)

## ANNEXE 5 : ATELIERS PROSPECTIFS - MÉTHODES

### Atelier "Scénarios extrêmes"

Un scénario extrême est un scénario qui pousse de façon radicale un paramètre précis, en laissant les autres plus ou moins en l'état. Par exemple, la disparition totale de l'humain dans les interactions. Ou un scénario qui retourne radicalement une proposition considérée comme acquise : par exemple, l'assistant vocal sans usage du langage (que des sons) ; un assistant vocal psychanalyste qui écoute mais ne parle pas...

Un scénario extrême permet de se projeter et de libérer les imaginaires.

Les scénarios extrêmes traités en atelier :

### > Scénario 1 : Cacophonie dans le foyer

Tous les objets de la maison se commandent par la voix. L'utilisateur interagit avec chacun selon les spécificités du modèle. Parfois, plusieurs objets se déclenchent en même temps et l'utilisateur s'emmêle les pinceaux avec les commandes d'activation et se perd dans les différents modes d'interaction.



### > Scénario 2 : Ces voix dans ma tête

Dans un monde où le numérique aurait réussi à pénétrer les remparts du cerveau, Me<sup>2</sup> est une prothèse mentale qui permet d'accompagner les porteurs selon leurs besoins propres : aide à la décision, coaching intérieur, aide à la communication orale grâce à la vocalisation des pensées pour les personnes atteintes de surdit  ou de probl mes phonatoires.

### > Sc nario 3 : Teacher bot

Un assistant vocal disponible sur smartphone est mis   disposition des coll giens. Cet assistant est destin  au soutien scolaire en dehors des heures de classe et ne s'active que dans une plage horaire d finie. Cet assistant a pour objectif de favoriser l'autonomie du coll gien dans ses r visions, dans l'aide aux devoirs, dans l'acquisition de connaissances g n ralistes - ma trise de la langue - ou de comp tences sociales telles que la prise de parole en public.

### > Sc nario 4 : La tour de Babel 2

“Les murs ont des oreilles” : la soci t  de surveillance s'est impos e, l' coute est partout et la voix est l'interface de pr dilection. Chacun   son niveau - cercle familial, cercle professionnel, soci t  civile, activistes - d veloppe des strat gies d' vitement. De nou-

veaux langages (sons, onomatop es, gestes) apparaissent et se renouvellent en permanence pour tromper la compr hension des machines.

### > Sc nario 5 : L'assistant vocal, porte parole des “sans voix”

Une  quipe de scientifiques s'est saisie de la technologie vocale pour donner de la voix aux plantes et aux animaux. Le bot Plante a un r le de confident, aupr s duquel les usagers peuvent d poser leurs pens es et r flexions et il est  galement possible d'activer le mode discussion afin d' changer avec lui.

### > Sc nario 6 : Le premier Syndicat vocal “UnionBot”

UnionBot voit le jour apr s une gr ve majeure dans la Silicon Valley. Il est la voix et l'oreille des salari s : UnionBot r colte l'ensemble de leurs questions et les restitue en assembl e g n rale pour les mettre en discussion, pour identifier les chantiers prioritaires et pour n gocier. UnionBot si ge au conseil d'administration : il enregistre, il restitue, il met en d bat.

### > Sc nario 7 : Action sociale vocale

Le minist re de la sant  met   disposition des AV pour palier au manque de personnel, en particulier dans le domaine de la sant 

(h pitaux, Ehpad, aide   domicile ...). Les AV garantissent une compagnie et une  coute gratuites automatis es, compensant la technicisation du soin et pour apporter la dimension relationnelle manquante. Un op rateur priv  est mandat  par le minist re de la sant  pour fournir ces AV dont le contenu est enrichi par les diff rents acteurs (mutuelles, s curit  sociale, associations...).

- » M thodologie de l'atelier : [fing.org/IMG/pdf/Deroule.pdf](http://fing.org/IMG/pdf/Deroule.pdf)
- » Les cartes “Profil” : [fing.org/IMG/pdf/Profils.pdf](http://fing.org/IMG/pdf/Profils.pdf)
- » L'ensemble des sc narios extr mes : [fing.org/IMG/pdf/Scenarios\\_Extremes\\_HyperVoix.pdf](http://fing.org/IMG/pdf/Scenarios_Extremes_HyperVoix.pdf)

## ANNEXE 6: LE DESIGN FICTION

### 1. LE DESIGN FICTION POUR EXPLORER LES FUTURS ALTERNATIFS ET FRICTIONNELS DES AV

#### > Le Design Fiction, posture exploratoire et critique du design

*Le Design Fiction est une approche qui détourne les codes et les outils du design afin “d’inspirer de nouveaux imaginaires liés au futur d’une manière non prédictive (“le futur sera ainsi”) et non prescriptive (“le futur doit être ainsi”). Il s’agit d’imaginer des scénarios (étape de fiction) puis de les incarner dans des objets et services fictionnels (étape de design) pour contribuer à ouvrir de nouvelles perspectives et à les mettre en débat.” (formulation de Design Friction).*

#### > Le Design Fiction pour la médiation inclusive des futurs

Parce qu’il permet à un public de se projeter dans des futurs par le prisme du quotidien, “comme s’ils étaient déjà là”, le Design Fiction constitue une forme de médiation qui vient enrichir la discussion multipartite

autour de possibles prospectifs, pour finalement prendre des décisions concertées et mieux éclairées au présent. Cette approche offre une porte d’accès bienvenue à des problématiques émergentes (incertitude), systémiques (multifacteurs, multi-impacts) et structurelles (redéfinition des règles du jeu d’une industrie).

#### > Le Design Fiction pour aider la prise de décision

Elle se trouve donc toute indiquée pour accompagner l’exploration des enjeux, impensés, angles-morts et points de frictions induits par l’émergence des technologies vocales et conversationnelles. S’intéresser aux implications de ces usages naissants et à la réalité des visions portées par ces objets et services en amont de leur mise sur le marché est une étape cruciale dans le processus d’innovation, d’un point de vue éthique et commercial.

### 2. L’APPORT DU DESIGN FICTION DANS LE PROJET HYPERVOIX

#### > Convier des publics variés à la réflexion et à la discussion

Soucieux d’impliquer des profils variés de

contributeurs, experts (entreprises partenaires, chercheur·ses, concepteurs·trices, journalistes) comme non-experts (futur.es utilisateur·trices, étudiant.es, grand public) de la thématique des interfaces vocales, pour explorer de manière inclusive des problématiques qui nous concernent tous,

#### > Le Design Fiction pour démystifier l’attendu et accepter l’impensé

Construits sur la base des scénarios extrêmes imaginés au Temps 2, les scénarios de design fiction visent aussi bien à défricher et évacuer les réflexes de pensée et les clichés dans les évocations des interfaces du futur qu’à libérer des imaginaires nouveaux et alternatifs. A l’écart des visions techno-naïves ou des dystopies faciles, dans un entre-deux volontairement gris, frictionnel, ils mettent en tension des implications socio-économiques, culturelles et politiques associées à ces innovations.

#### > Des signaux faibles aux scénarios

La matière de l’étape de recherche (détection de signaux faibles, identification de points de tension, prise en compte de possibles radicaux) est mobilisée pour imaginer des quotidiens au futur. Ces scénarios d’anticipation racontent les modes de vie, les joies, les questionnements, les relations,

les problèmes, etc. des habitants du futur proche. Il s'agit de dépeindre un monde alternatif, mais toujours crédible et plausible, et donc en tenant compte de sa complexité, un monde inter-relationnel dans lequel s'affrontent les visions et les intérêts d'entités variées. Il s'agit davantage de raconter un monde qu'une histoire - le fameux «tell worlds rather than stories» de l'auteur Bruce Sterling : (faire) se projeter dans une sélection de facettes d'un monde fictionnel (world building), tout en laissant de la place à l'imagination du public pour se développer dans les zones de non-dits.

### 3. LES DESIGN FICTIONS DU PROJET HYPERVOIX

Parmi l'éventail de productions ayant émergé de l'atelier de "Scénarios extrêmes", les équipes Hypervoix et Postillon-Propective ont co-sélectionné quatre scénarios à être remaniés puis développés en design fictions. Retrouvez-les ci-dessous, ainsi que les éléments de discussion qu'ils ont permis de faire émerger lors du temps de débat.

#### A. Leaderbot

Et si un AV permettait de centraliser les re-

quêtes et les commandes d'objets de la maison ?

Et si un assistant vocal devenait notre nouveau colocataire ? Le "bot leader" est un assistant généraliste, non spécialisé. Il partage des informations en interaction avec les individus mais aussi avec les différents systèmes techniques. C'est un dispositif anti-cacophonie qui gère les priorités, la hiérarchie des objets afin de limiter les notifications/interactions vocales inopportunes. Le "bot leader" est paramétrable, proposant différents modes utilisateur-riche-s, selon différents profils afin de gérer les identités individuelles ou collectives et les accès. Ce dispositif peut privilégier des notifications écrites ou visuelles en fonction du contexte. Le "bot leader", conçu comme un major-dome numérique discret, est une réponse envisageable à l'image des hubs domotiques proposés aujourd'hui par les acteurs technologiques, alliant interactions vocales, commandes tactiles et écran de contrôle.

[> Découvrez la bande sonore de la design fiction](#)

#### Le scénario du Leaderbot : futur préférable ou futur indésirable ?

Ce futur est une perspective préférable car...

- » "Leaderbot" me permet de déléguer les tâches du quotidien qui m'ennuient ou ne présentent pas d'intérêt, comme dresser une liste de course, afin de me dégager du temps pour faire les choses qui m'intéressent vraiment, comme me rendre à une exposition ou pouvoir discuter avec mes amis.
- » "Leaderbot" est présent mais il n'est pas au centre des interactions ; on n'est pas obligé d'en passer par lui pour discuter avec ses voisins. Il est comme un nouvel entrant dans l'habitat, qui rend service à tout le monde.
- » "Leaderbot" présente un mode d'interaction ultra-personnalisé et donc satisfaisant.
- » "Leaderbot" m'est sympathique, à condition qu'il soit possible de le débrancher quand je le souhaite.

Ce futur est une perspective indésirable car...

- » "Leaderbot" endosse un rôle central de médiateur, cela pourrait entraîner le risque de perdre en capacité à se coordonner et s'organiser.
- » "Leaderbot" risque de faire de nous des assistés. La concertation est un effort qui demande à s'apprendre et peut se désapprendre.
- » Ici l'efficacité est priorisée par rapport à la concertation.

- » Les données captées et gardées en mémoire ne semblent pas être en mon contrôle : sont-elles placées sur des serveurs sans mon consentement et accessible à des tiers ?

## B. BVP

Et si un assistant vocal ouvrait de nouveaux modes d'apprentissage ?

Dans ce scénario, le “Ministère de l'Éducation Nationale et des outils techno-pédagogiques” ambitionne de démocratiser le soutien scolaire en fournissant à chaque élève, de la 6ème à la terminale, un “BVP” ou Boîtier Vocal Pédagogique. Cet assistant vocal a pour fonction d'écouter ce qui se passe en classe et d'interagir avec l'élève hors des heures de cours pour le faire réviser de manière personnalisée, l'aider à creuser des sujets et à mémoriser, le faire se questionner par rapport à ce qui lui a été enseigné (consolidation de l'esprit critique). En faisant le pari du vocal, l'idée est de fournir un outil pédagogique qui ne surcharge pas le champ visuel des élèves avec un écran supplémentaire et aussi de pouvoir détecter précocement les cas de souffrance ou harcèlement scolaire, grâce à l'analyse émotionnelle et lexicale.

Et si on imaginait un mode d'interaction

maïeutique, où l'assistant, tel un Socrate des temps modernes, pourrait poser des questions au lieu de formuler systématiquement des réponses ? Car s'éduquer ce n'est pas seulement acquérir de la connaissance mais aussi développer des réflexions propres, remettre en cause ses préjugés, mûrir son esprit critique. Il faut “apprendre à apprendre” comme l'a formulé l'une des participantes de l'atelier. Ce pourrait être aussi la stratégie déployée à travers ce type d'assistant vocal : sortir du “par coeur” pour faire comprendre et mémoriser par connexions d'idées, astuces mnémotechniques, etc. Sans oublier la capacité potentielle, grâce à l'apprentissage machine, de mieux cerner les spécificités de chaque élève pour adapter les techniques d'apprentissage à ses besoins, complétant ainsi le mode d'apprentissage nécessairement normé du système scolaire.

[> Découvrez la bande sonore de la design fiction](#)

### Le scénario du BVP : futur préférable ou futur indésirable ?

- » Ce futur est une perspective préférable car...
- » Le “BVP” peut compenser les inégalités au niveau de l'accompagnement du tra-

vail à domicile des enfants.

- » Le “BVP” peut être intéressant pour les étudiant.es qui doivent préparer des concours (ex : médecine).
- » Le “BVP” serait forcément gratuit si c'est un service poussé par le secteur public.
- » Les modules payants pourront certainement être “craqués” facilement, à l'instar de la plupart des logiciels existants.
- » Si une partie du service est payant, il faudra penser des aides financières pour que cela ne crée pas des inégalités.
- » La personnalisation de l'accompagnement ne crée pas nécessairement de dépendance ni d'exclusivité. Par exemple l'usage de Spotify n'empêche pas d'aller à des concerts.
- » L'assistant peut relancer l'élève (push) mais rien n'empêche que l'élève soit proactif et que le bot stimule sa curiosité (pull).
- » Le “BVP” permet aux personnes timides de la classe, qui n'osent pas poser des questions, d'obtenir des réponses et compléments d'information.

Ce futur est une perspective indésirable car...

- » L'assistant est trop présent par rapport à la parole laissée à l'élève.
- » Inadapté pour des enfants trop jeunes

car c'est le moment où il faut aussi "apprendre à apprendre".

- » Le "BVP" donne des réponses immédiates et n'encourage pas le questionnement : il faut que le boîtier permette d'améliorer l'intelligence et la capacité à réfléchir plutôt que donner les réponses tout de suite.
- » Ici, c'est l'école privée portative. S'il faut payer à chaque chapitre on va augmenter les inégalités du système scolaire qui sont déjà conséquentes.
- » Pourquoi aller à l'école si on peut tout apprendre avec son "BVP" depuis chez soi ? Il y a un risque de disparition des espaces de vie et d'apprentissage collectifs ; si on délègue tout à la machine, c'est un échec.
- » Le "BVP" risque de créer une forme de dépendance intellectuelle avec une machine. Il ne favorise pas beaucoup les échanges et les interactions entre les étudiants.
- » Le "BVP" analyse les sentiments de l'élève avec détection de signaux témoins de la condition psychologiques sans prendre soin de proposer un cadre de confiance adéquat.
- » Le fait qu'on puisse donner un prénom au bot c'est qu'il y a quelque chose de très relationnel dans la conception de l'interaction. Ceci ajouté à la détection de signaux faibles psychologiques, une pré-

sence forte et la personnalisation poussée du bot : ça va plus loin que la simple assistance aux devoirs et c'est un problème.

- » Pour quels élèves cet outil est-il pensé ? Le scénario montre un niveau de langage et un échange très fluide qui masque des situations potentiellement plus difficiles - par exemple où il faudrait cinq minutes à un élève pour formuler une réponse ou une question. Combien de personnalités faudrait-il penser pour que le "BVP" puisse s'adapter à tous les élèves ?
- » C'est une déshumanisation de l'apprentissage.
- » Il y a un risque de dépendance ; il faudrait plutôt favoriser l'autonomie des élèves.
- » Chercher par soi-même est un aspect important de l'apprentissage.
- » Ce n'est pas un divertissement mais une technologie d'apprentissage donc si il y a une dépendance ça n'a pas les mêmes conséquences.
- » L'usage de ce type d'outil est directement lié au niveau d'autonomie de l'apprenant.e : il faudrait d'abord déceler par l'humain si la personne pourrait projeter un parcours d'apprentissage sur la machine.

## C. Arlette

Et si un assistant vocal était dédié à l'action militante collective ?

Dans ce scénario, l'assistant vocal propose une zone d'écoute (à priori) sécurisée, localisée (dans l'habitacle d'un VTC), qui invite à la confiance, à la prise d'information capacitante et surtout, aide à effectuer un tri dans ce qui lui est confié, pour faire remonter une sélection de revendications destinées à impacter le système en place. La design fiction "Arlette" invite à penser une technologie vocale qui serait le fruit d'un volontarisme activiste et d'un développement indépendant, loin du contrôle des GAFAM. L'usage de l'assistant est ouvertement politique et dépasse l'individu dans le sens où la finalité se veut l'action collective et le progrès social.

"Arlette" donne l'illusion jouissive d'une technologie neutre au sens où elle répond par de l'information factuelle à des questions précises et qui relèvent du domaine du juridique. Peut-être est-ce là la clé d'une interaction saine au sens de conscientisée : se poser la question de ce qu'on peut se permettre de demander à la machine - et a contrario de ce qu'il vaut mieux éviter de demander afin de minimiser les risques (réponses fortement biaisées, maladroitement, hors sujet, erro-

nées, discriminantes, etc.). Alors, peut-être faut-il envisager de concevoir des assistants contextualisés, qui seront rendus volontairement incapables de répondre à toutes nos questions ou d'échanger sur tous les sujets. Une sorte de geofencing<sup>98</sup> par champ de compétence. La piste d'un assistant vocal très spécialisé (e.g. juriste) semble promettre un meilleur degré de fiabilité et une meilleure lisibilité quant aux limites de son rôle. C'est peut-être aussi là la réponse au risque grandissant des fake news : des sources d'informations restreintes mais vérifiées à priori, au lieu d'un fact-checking à posteriori.

### > Découvrez la bande sonore de la design fiction

Le scénario d'Arlette : futur préférable ou futur indésirable ?

Ce futur est une perspective préférable car...

- » "Arlette" est intégré à la routine et à l'espace de travail, permettant de travailler tout en agissant dans le même temps pour une amélioration de ses conditions de travail. Un avantage précieux là où militer demanderait un temps d'investissement qui viendrait s'ajouter diffici-

lement à des horaires déjà lourdes.

- » "Arlette" permet à cette communauté de travailleurs de mieux protéger leurs droits et devenant mieux informés grâce au sourcing d'information en ligne.
- » Rebond : et si chacun pouvait poser des questions sur tout type de contrat (de travail, d'assurance, RGPD, etc.) : "Suis-je protégé?" "Est-ce que je paye le bon prix pour tel ou tel service?". Et si on pensait un "assistant juriste" qui permettrait d'avoir accès à des infos complexes de façon simple et neutre ?
- » L'oral / le vocal est une bonne aide dans ce contexte où il évite de devoir lire des documents pour creuser un sujet.
- » "Arlette" est un assistant vocal qui réussit à connecter les humains et même à les coordonner. Ces nouveaux modes d'actions peuvent être très utiles dans un contexte persistant d'uberisation ou de syndicalisme qui n'aurait pas évolué (inaction, faible impact, absence des interlocuteurs).
- » "Arlette" est intéressant de par sa neutralité apparente. Cela pourrait être utile lors des actions sociales, comme les manifestations, pour opérer un décompte transparents des personnes mobilisées. Ce pourrait aussi être utile pour imaginer une alternative au corporatisme et à l'affiliation politique des travailleurs.
- » La design fiction est une anticipation de

ce qui va se produire tôt ou tard : l'action collective permettra d'imposer d'autres modalités aux plateformes qui aujourd'hui jouent selon les règles qu'elles ont imposées.

Ce futur est une perspective indésirable car...

- » "Arlette" amène à ne pas sortir de son véhicule, à rester un maximum de temps dans sa bulle de travail, ce qui renforce le côté aliénant du métier et la course à la productivité - on ne s'arrête même plus pour faire une réunion syndicale.
- » "Arlette" pourrait servir à créer du lien présentiel au lieu de renforcer l'isolement individuel dans l'habitacle.
- » "Arlette" présente un risque de confusion entre l'humain et la machine dans l'interaction : on entend presque une amie qui comble le vide relationnel de la journée de travail (en ce sens, "Arlette" n'est d'ailleurs pas très réaliste).
- » La design fiction présente une interaction avec une chauffeuse très calme et intelligible. Si l'échange avait été téléphonique entre deux humains, il aurait certainement été moins apaisé. Cet échange "neutre" manque de réalisme.
- » Il est illusoire de penser qu'"Arlette" permet la coordination d'actions contestataires discrètes. A ce moment là, quel mode de cryptage ? Quel cadre de confiance ? Qui dispose des données ?

98 Le geofencing ou géopérage est une technologie permettant de définir des barrières virtuelles pour un appareil électronique.

## D. Végétalk

Et si on imaginait un assistant vocal pour l'activisme écologique ?

Et si le vocal oeuvrait pour la "République du vivant" ? Des assistants vocaux pour sensibiliser aux enjeux environnementaux ? L'idée peut sembler paradoxale, du moins provocatrice, quand on sait qu'une technologie vocale, notamment lorsqu'elle embarque une forme "d'intelligence" est loin d'être sans impact sur l'environnement. Mais pouvoir donner de la voix à des objets, c'est aussi pouvoir donner la parole au vivant muet. Et si la nature prenait la parole pour assurer sa propre défense ? C'est ce que simule la technologie présentée dans la design fiction "Végétalk", dans un but de sensibilisation et de protection du patrimoine environnemental.

Dans la lignée du "Parlement des choses" de Bruno Latour, ou dans les pas de l'écrivain et journaliste Aymeric Caron, antispéciste - idéologie qui consiste à placer les intérêts des animaux au même niveau que ceux des humains - cette design fiction propose un futur où l'on s'avance (peut-être) vers une écoute et une défense de la parole végétale. La "République du vivant" (formule d'Aymeric Caron) ainsi constituée serait une "bio-

démocratie" où les arbres du poumon vert d'Amazonie, pour commencer, se verraient garantir leur préservation et la défense de leurs intérêts - à noter que la justice de Colombie a déjà reconnu la forêt amazonienne comme sujet de droit en avril 2018.

[> Découvrez la bande sonore de la design fiction](#)

### Le scénario Végétalk : futur préférable ou futur indésirable ?

Ce futur est une perspective préférable car...

- » Le dispositif présenté dans "Vegetalk" est un bon moyen de sécurité parmi d'autres, mais présente un risque d'habitude qui le rendrait inopérant.
- » La voix comme système d'alerte contre les agressions peut être une super idée.
- » Aujourd'hui les machines n'ont pas le droit de parler "pour" les humains, mais ce peut être une très bonne manière de stimuler la prise de conscience par rapport au vivant : c'est un instrument qui pourrait changer les imaginaires.

Ce futur est une perspective indésirable car...

- » L'argument écologiste ne tient pas : l'installation ne pourrait pas être déployée sur tous les arbres, et le jeu en vaut peut-être la chandelle.

- » On a plutôt envie de se servir de ce genre d'interface pour donner à apprendre plutôt que pour défendre.
- » La cible n'est pas adaptée : pourquoi pas pour s'adresser aux enfants, comme outil pédagogique ? En tant que tel c'est un énorme dispositif pour une solution à très court terme.
- » C'est paradoxal d'introduire un objet connecté dans la nature, ayant un impact écologique négatif : une vraie fausse bonne idée.
- » Une nuisance sonore qui peut impacter les autres espèces.
- » La tendance qui consiste à faire parler les plantes et les animaux avec un langage humain est discutable : pas sûr que cela sensibilise à ces être vivants et que ça permette de mieux les comprendre, vu qu'on plaque de l'humain sur du non-humain.
- » L'idée est intéressante, mais constitue une forme d'instrumentalisation de la nature et du vivant au profit d'intérêts beaucoup plus catégoriels que "Dame Nature".
- » Ce n'est pas parce que ça parle qu'on suscite plus d'empathie. Cette vision relève de l'utopie ; le dispositif serait dissuasif dans un premier temps, mais ensuite il faudrait jouer sur d'autres leviers plus pédagogiques.

- » Est-ce vraiment à ceux qui coupent les arbres qu'il faut d'adresser ? Qui sont les bons interlocuteurs ? Les braconniers ont leurs raisons et en arrêtant ces gens-là on ne va pas leur apprendre grand chose : comment démanteler le système plutôt que ceux qui en sont victimes ?
- » Ce futur est anxiogène. Mettre des capteurs dans la forêt amazonienne pour détecter des humains, c'est nier que le problème est "humain", et le traiter de façon inquiétante.
- » La design fiction mentionne "l'implication des tribus locales", mais qu'en est-il exactement ? Un tel dispositif est-il en accord avec leur culture ? Un discours vendeur et démagogique ?
- » C'est un nouvel outil de contrôle qui fait peur et pousse encore plus loin la perspective d'un état de surveillance, sans droit à l'erreur.
- » Imaginer ce dispositif dans un réseau urbain est une perspective indésirable : on va davantage entendre les poteaux parler que les gens ; une pollution sonore urbaine (saturation, cacophonie).