

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326176795>

Conception d'interactions éthiques et durables entre l'humain et les systèmes d'intelligence artificielle. Le cas de l'expérience vécue des usagers de l'IA vocale

Article in *Revue D Intelligence Artificielle* · December 2018

DOI: 10.3166/RIA.32.417-445

CITATION

1

READS

1,140

2 authors:



Myriam Fréjus

Électricité de France (EDF)

36 PUBLICATIONS 194 CITATIONS

[SEE PROFILE](#)



Dounia Lahoual

Électricité de France (EDF)

17 PUBLICATIONS 39 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Smart Home [View project](#)



Human AI Interaction [View project](#)

Conception d'interactions éthiques et durables entre l'humain et les systèmes d'intelligence artificielle

Le cas de l'expérience vécue des usagers de l'IA vocale

Dounia Lahoual, Myriam Fréjus

EDF R&D - 7 Boulevard Gaspard Monge, 91120 Palaiseau, France

dounia.lahoual@gmail.com

myriam.frejus@edf.fr

RÉSUMÉ. Cette recherche vise à construire une vision de l'éthique située et systémique articulée à la notion de durabilité, afin de privilégier la conception d'interactions éthiques et durables entre l'humain et les systèmes d'intelligence artificielle (IA). Envisagée dans une logique d'aide à l'activité humaine, la conception éthique et durable de l'interaction homme-IA considère les contextes d'utilisation, les points de vue intrinsèques des utilisateurs et le développement de l'humain. Elle s'appuie sur un ensemble de critères de conception qui conditionnent l'acceptabilité et l'appropriation d'un système d'IA. Nous illustrons ces problématiques à partir de retours d'expériences d'usages de technologies vocales dans les sphères domestiques et professionnelles ; ceux-ci sont mis en regard avec les critères éthiques soulevés par les politiques publiques et les comités éthiques que nous synthétisons. Nous discutons ces éléments dans une logique d'éthique située, pour alimenter notre réflexion sur les interactions éthiques et durables à encourager dès la conception.

ABSTRACT. This study aims at building a systemic and situated ethical vision, around the notion of sustainability, in order to favor more ethical and sustainable interactions between humans and Artificial Intelligence (AI). We consider that the design of the human-AI interaction has to be based on the contexts of use, the subjects' intrinsic points of view as well as the development of the human being. To assist the human activity in an acceptable and appropriated way, the design of the AI based system must observe a set of design criteria. We illustrate these issues with a literature review on ethical criteria being discussed by public policies and ethics committees. We reflect on these results with a situated ethics point of view by considering the real uses of voice technologies in domestic and professional settings. This knowledge help to precise how to promote ethical and sustainable interactions during the design process of AI systems.

MOTS-CLÉS. Éthique, durabilité, interactions homme-système, intelligence artificielle, technologies vocales, assistants vocaux, activité domestique, usages, acceptabilité, appropriation, ergonomie, développement, utilisateurs.

KEYWORDS: ethics, sustainability, human-system interactions, artificial intelligence, voice technologies, voice assistants, domestic activities, uses, acceptability, appropriation, ergonomics, development, users.

DOI:10.3166/RIA.32.417-445 © 2018 Lavoisier

1. Introduction

1.1. Un contexte propice à l'émergence d'enjeux éthiques

Ces dernières années ont été marquées par de nombreux progrès dans le domaine de l'intelligence artificielle, et cela dans plusieurs secteurs : robotique, véhicules autonomes, traitement de la parole et compréhension du langage naturel, maisons intelligentes, etc. Parmi les récentes réalisations relayées dans les sphères médiatiques et scientifiques, nous retrouvons par exemple la victoire du programme informatique Alpha Go de Google contre le champion du monde de jeu de Go, Lee Sedol, en une série de cinq matchs en mars 2016. L'IA a également permis la description automatique du contenu d'une image, le développement de systèmes de reconnaissance faciale, de traduction instantanée, le déploiement des véhicules autonomes, etc. D'une part, ces nombreux succès ont été obtenus grâce au développement de l'IA, domaine de recherche interdisciplinaire orienté vers « *la compréhension, la modélisation et la réplique de l'intelligence et des processus cognitifs faisant appel à divers principes et appareils informatiques, mathématiques, logiques, mécaniques ou même biologiques* » (Frankish et Ramsey, 2014). Cette forme d'intelligence artificielle est d'autant plus accrue lorsque les systèmes sont dotés d'une autonomie leur permettant de percevoir les changements de leur environnement et d'adapter leur fonctionnement pour y exécuter des tâches complexes (Comest, 2017). D'autre part, ces avancées fulgurantes ont également été réalisées avec le déploiement de l'apprentissage automatique. Ce dernier permet de concevoir des programmes permettant à une machine de construire de nouvelles connaissances à partir de son expérience. L'homme entraîne ainsi le système en lui fournissant des données qui lui permettent d'apprendre et de déterminer les opérations à effectuer pour des tâches hautement complexes (CNIL, 2017-a ; Russel et Norvig, 2009 ; Heudin, 2017). Ces algorithmes apprenants possèdent la particularité d'être mouvants et imprévisibles, tout en nécessitant une immense quantité de données diversifiées nécessaire pour l'entraîner.

Ces éléments suscitent de nouveaux enjeux éthiques, tels que la question de la responsabilité des actions exécutées par les systèmes d'IA. En effet, les systèmes peuvent être conçus de manière à fonctionner indépendamment d'un utilisateur ou d'une autre machine dans un environnement complexe et variable. Leur autonomie et leurs capacités décisionnelles leur permettent d'« *interpréter les informations issues de ses capteurs, déterminer les actions les plus pertinentes en fonction de cette interprétation et de ses connaissances préalables, et calculer quand et avec quelles ressources ces actions seront effectuées* » (Allistene, 2014). Ces capacités questionnent notamment d'autres enjeux éthiques : le partage de l'autorité, le

pouvoir décisionnel, la capacité de reprise en main par l'utilisateur, la pertinence et la fiabilité des algorithmes qui fondent les décisions prises par le système, ou encore, l'attribution de la responsabilité en cas de dysfonctionnement et de préjudice causé par le système d'IA auprès des utilisateurs (Allistene, 2014 ; Comest, 2017 ; Villani *et al.*, 2018). Par ailleurs, le développement constant des mégadonnées analysées par les algorithmes d'auto-apprentissage est indispensable dans le déploiement de l'IA. Toutefois la commercialisation et le traitement de ces données à caractère personnel, collectées à partir des médias sociaux ou des dispositifs connectés à internet (« l'internet des objets »), peuvent engendrer un ensemble de conséquences sur la vie privée des utilisateurs (exemple : révéler des informations sur les activités, comportements ou déplacements des personnes). Les algorithmes de recommandation représentent également une autre application de l'IA visant la recommandation d'un produit ou d'une offre de façon personnalisée en élaborant des prévisions sur les comportements et les préférences pour devancer les besoins ou orienter les choix des consommateurs. Ce filtrage d'informations serait ainsi susceptible d'augmenter *la propension des individus à ne fréquenter que des objets, des personnes, des opinions, des cultures conformes à leurs propres goûts et à rejeter l'inconnu*, allant ainsi à l'encontre de l'exposition à la diversité culturelle (CNIL, 2017-a).

1.2. De l'éthique à la durabilité : des critères pour concevoir des interactions entre l'homme et les systèmes d'IA

Les considérations éthiques précédemment évoquées démontrent l'importance de produire une réflexion sur les bouleversements technologiques et leurs impacts sur la société et les usagers, et cela « *dès lors que l'évolution technologique déplace la limite entre le possible et l'impossible et nécessite de redéfinir la limite entre le souhaitable et le non souhaitable* » (CNIL, 2017-a ; Villani *et al.*, 2018). En effet, la dimension éthique s'impose notamment lorsqu'il s'agit d'alimenter la réflexion sur l'acceptabilité des technologies à risques, générant des incertitudes et des conflits de valeurs (Kermisch, 2014 ; 2016). Dans ce contexte, nous voyons émerger un ensemble de comités éthiques et de rapports commandés par les législateurs afin de fournir des principes et des lignes directrices éthiques visant à faire de l'IA un enjeu maîtrisable, au service de l'homme et démystifiée pour une plus grande transparence, sûreté et acceptabilité sociale (Rapport parlementaire, 2017). Au service de l'humain et de l'environnement, ces principes ont été proposés pour accompagner les concepteurs dans le développement et l'intégration de ces nouveaux systèmes autonomes et intelligents dans notre société. Et ceci, afin qu'ils préservent les droits de l'homme, le bien-être humain, et qu'ils assurent une responsabilité en cas de préjudice, une transparence et une sensibilisation aux risques d'utilisation abusive (IEEE Global Initiative, 2017).

Certains outils ont également été proposés tels que les « matrices éthiques » afin d'éclairer le débat citoyen et la prise de décision au niveau politique et industriel. Ces matrices consistent à dégager des enjeux éthiques associés à une technologie par la combinaison de prémisses morales (telles que les principes d'autonomie, de

bienfaisance ou de justice, fournissant un cadre à la réflexion éthique) associées à une démarche participative en convoquant les acteurs potentiellement concernés (Mepham, 2000). Bien que ce type de méthode promeuve une transparence du processus décisionnel et la prise en compte d'une multiplicité de points de vue pour un cas étudié, les matrices éthiques portent diverses contraintes : le temps consacré à la discussion, la quantité de participants à mobiliser, l'impartialité lors de la construction de la matrice est confiée au preneur de décisions et l'aspect réducteur de la matrice si elle est prédéterminée sans la mise en place d'un débat participatif (Kermisch, 2014).

Les travaux provenant de la philosophie et de la roboéthique nous permettent de délimiter la notion d'éthique en ce sens : l'éthique correspondrait à une réflexion argumentée sur les comportements à adopter en vue du bien agir. Cette réflexion viserait la prise de décisions en situations concrètes et complexes (Catellani *et al.*, 2017) en s'appuyant sur des principes en faveur des droits humains tels que l'autonomie (les rendre libres et capables), la bienfaisance (contribuer au bien-être des humains et ne pas leur porter volontairement préjudice), ou encore la justice (Kermisch, 2014). Au fondement de la morale, elle suppose un choix personnel et difficile à réaliser. L'éthique n'est pas toujours référé à des normes, mais plutôt une réflexion subjective qui précède l'action (Lagarrigue et Lebe, 1997). Cette sagesse pratique se cristallise dans « *une intelligence circonstanciée, en contexte pour rendre l'acte éthique le plus durable possible* » (Fleury, 2017). De plus, l'éthique n'est guère fixée une fois pour toutes, elle tend à s'actualiser et à faire l'objet de réflexion et de discussion au sein de la communauté sociale. Elle évolue donc de façon entrelacée avec les progrès technologiques, n'ayant cessé de remettre en question les principes et les valeurs humaines.

Pour ces raisons, cette vision de l'éthique en action nous invite à privilégier une vision systémique, située et finalisée des acteurs, de leurs situations et de leurs motivations associées à l'organisation sociale et aux systèmes socio-techniques impliqués. Une meilleure compréhension et intégration de l'éthique dès la conception, deviendrait ainsi un moyen pour soutenir les interactions entre l'homme et les systèmes d'IA. Associée à la prudence en action (Morel, 2017) et au bénéfice de l'humain, l'éthique considère le bien-être et le bien-agir au même titre que la durabilité telle qu'elle a pu être discutée en ergonomie. Ancrée dans les travaux relatifs au développement durable, l'interaction durable se définit comme une dynamique et un équilibre durable entre systèmes, « *impliquant l'usage et la connaissance appropriée de technologies, de façon à empêcher que l'humain et l'environnement puissent subir des dommages ou être exposés au moindre danger* » (Garcia-Acosta *et al.*, 2012). Nous positionnons notre vision de la durabilité comme un critère de conception favorisant une interaction efficace entre l'acteur et son environnement perdurant sur le temps : « *Par situation durable, nous entendons une situation résultant d'une appropriation de nouvelles ressources (...) incorporées à l'activité (...) Durable renvoie à la dimension temporelle et évolutive du couplage. La qualité du couplage se doit de perdurer alors que chaque élément autonome la constituant à sa propre évolution. Il s'agit donc plus de viser un processus se traduisant par un bien-être des acteurs allant au-delà de l'efficacité de l'action*

domestique » (Fréjus, à paraître). En intégrant la notion de durabilité dans notre démarche, nous nous orientons ainsi vers une logique d'aide à l'activité humaine et de transformation de la situation (*op. cit.*). Ainsi, comme l'éthique, la durabilité contextualise la relation humain-système et ancre les systèmes dans une logique de bien-être au service de l'homme. En complément, elle invite à considérer une dimension temporelle et développementale dans cette relation.

Pour atteindre une situation durable intégrant des systèmes d'IA, il est indispensable de privilégier les critères d'utilité, d'utilisabilité, d'acceptabilité et d'appropriabilité qui favorisent, entre autres, le développement humain continu impliquant l'évolution des usages, un apprentissage, une réflexivité sur les activités et leurs conséquences, etc. (Fréjus, à paraître). Ces critères sont au cœur de la démarche de conception et conditionnent l'adoption des systèmes sur le long terme. L'utilité renvoie à l'adéquation du dispositif et de ses services aux besoins et aux attentes des utilisateurs (Shackel, 1991). Ce critère tend le plus souvent à s'articuler à celui de l'utilisabilité qui vise à « *atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié* » d'après la norme ISO 9241-11 (1998). Quand bien même ces critères seraient intégrés dès la conception, le dispositif ne serait toujours pas assuré d'être accepté, ni approprié (Dubois et Bobillier-Chaumon, 2009), car l'activité incorpore un aspect dynamique et longitudinal imprévisible qu'il importe d'investiguer en situation réelle. Cette caractéristique est davantage mise en lumière avec la notion d'appropriation, et sa déclinaison en conception, l'appropriabilité (Haué, 2004). L'appropriation représente l'intégration, partielle ou totale, d'un objet, d'un outil ou d'un dispositif au corps propre ou la culture propre de l'acteur. Ce processus s'inscrit sur une perspective temporelle longue intégrant les histoires d'utilisations, les représentations des utilisateurs, et leurs nombreux épisodes de succès et d'échecs colorant ce parcours singulier (Pelletier et Moreau, 2008 ; Proulx, 2002). Cette dimension de l'appropriabilité renvoie ici à la capacité d'un dispositif à s'adapter et à réagir de façon pertinente selon le contexte, les préférences et le développement des utilisateurs sur le long terme (Dubois et Bobillier-Chaumon, 2009). C'est pourquoi les nouvelles perspectives d'autonomie des systèmes à base d'IA re-questionnent l'appropriabilité. Quant au critère d'acceptabilité, il renvoie aux raisons pour lesquelles un dispositif peut être susceptible d'être *en adéquation avec les pratiques, les ressources, les objectifs, le système de valeurs et les normes éthiques des utilisateurs pour être acceptable* (Dubois et Bobillier-Chaumon, 2009 ; Nielson, 1993).

1.3. Les technologies vocales¹ : de nouveaux défis questionnant les dimensions liées à l'éthique, l'acceptabilité et l'appropriation

Parmi les récents progrès de l'IA, nous retrouvons l'essor de la commande vocale avec l'émergence de robots conversationnels (chat bots) et les assistants

1. Ces technologies vocales permettent de combiner les systèmes de reconnaissance vocale à des méthodes de synthèse et de commande vocale (Minker et Néel, 2002). Elles incluent notamment les assistants vocaux tels qu'Alexa d'Amazon, Google Home, ou Siri d'Apple.

vocaux répondant aux requêtes vocales, tels que Siri d'Apple, Google assistant, Cortana de Microsoft et M de Facebook retrouvés communément dans les ordinateurs, les smartphones et les tablettes (INRIA, 2016 ; CNIL, 2017-a ; Guillaud, 2016). D'abord limités à l'utilisation du smartphone, les assistants vocaux intelligents intègrent les habitats avec l'arrivée des enceintes intelligentes telles qu'Alexa d'Amazon ou Google Home (Pestanes et Gautier, 2017). Un assistant vocal intelligent représente un logiciel intégré dans un dispositif tangible, disposant de la faculté de dialoguer avec un humain pour réaliser des tâches sur ordre vocal des utilisateurs (lancer des musiques, connaître l'actualité ou la météo, répondre à des questions à partir d'une recherche web, rappeler des événements liés à son agenda, piloter des objets connectés, etc.). Disposant d'une capacité d'apprentissage automatique, ces intelligences artificielles se développent continuellement sans l'intervention des développeurs, apprennent de leurs erreurs, et cela, tout en intégrant des informations sur les utilisateurs afin de proposer des services contextualisés et personnalisés (Lefevre, 2017 ; Edip, 2017).

Si la question de la bonne intégration de la reconnaissance de la parole dans les interfaces et produits grand public n'est pas nouvelle (Portolan *et al.*, 1999), la maturité technologique des assistants actuels questionne l'utilisation réelle qui peut en être faite tant par ses utilisateurs finaux que par les concepteurs, en lien avec les autres formes d'interaction. Nous considérons que pour développer ces nouveaux systèmes vocaux ou les intégrer à des produits ou services dans des logiques éthiques et durables, il importe d'étudier et de comprendre les situations d'utilisation dans lesquelles ils s'insèrent, ainsi que les leviers et les freins à leur appropriation. Intégrées dans un contexte de dispositifs ambiants, la commande et la reconnaissance vocale constituent ainsi de nouveaux défis en apportant de nouvelles situations et formes d'utilisation qui questionnent en profondeur les principes d'utilisabilité, d'appropriabilité et d'acceptabilité.

Les usages de l'intelligence artificielle développés dans de nombreux secteurs soulèvent aussi des problématiques d'éthiques variées : sécurité, respect de la vie privée, responsabilité, biomimétisme, reprise en main technique, justice et équité, diversité culturelle, etc.

Afin d'appréhender cette problématique de façon située et systémique, nous proposons dans la prochaine partie de mettre en perspective les points de vue des législateurs sur les questions éthiques avec ceux des usagers de l'IA appliquée aux technologies vocales, ces derniers étant documentés à partir de retours d'expériences d'interactions vécues en situations réelles. Dans une dernière partie, nous concluons et discuterons des résultats à partir de notre vision d'une éthique située et systémique. Cette dernière sera combinée à la notion de durabilité comme critère de conception pour les interactions homme-ia. Enfin, nous préciserons nos perspectives de recherche sur une étude en situation réelle de systèmes d'interaction vocale.

2. Critères et préconisations pour l'éthique : articuler les perspectives des législateurs à une vision contextuelle de l'éthique à partir des expériences vécues des usagers de l'IA

Cette contribution repose en premier lieu sur un état de la littérature intégrant des travaux scientifiques en bioéthique, roboéthique, philosophie, ergonomie, ingénierie des IHM, sciences juridiques, ainsi que des rapports d'un ensemble de commissions nationales et internationales de réflexion sur l'éthique, l'IA et la robotique dont certains ont été commandés par des législateurs. Ces discussions traitent des questions éthiques relatives à la conception et à l'usage des nouvelles technologies incorporant de l'IA et impactant les utilisateurs et les relations homme-machine. Parmi ces éléments, nous avons regroupé en catégories ces différents enjeux éthiques discutés par les politiques publiques, les comités éthiques et expérimentés par les utilisateurs : *protection de la vie privée et confidentialité, diversité culturelle et servicielle, justice et équité, autonomie et reprise en main, biomimétisme et liens affectifs*.

En second lieu, nous avons mené un état de l'art sur l'utilisation des technologies vocales en contexte domestique et professionnel. Ce travail s'appuie sur une diversité de sources, telles que des travaux scientifiques en design, en ingénierie IHM, en ergonomie, des enquêtes en marketing et en stratégie menées par des instituts spécialisés et des cabinets de conseils, ainsi qu'une diversité d'articles et de billets récents sur les pages du web². Puis, nous présentons ici les retours d'expérience soulevant des questions éthiques ainsi que des problématiques d'acceptabilité et d'appropriabilité liées à l'IA vocale.

Ainsi, dans les sous-parties suivantes, nous articulons les visions des législateurs et des utilisateurs pour chaque critère éthique présenté. Cette mise en perspective vise à identifier les points convergents et divergents entre une vision réglementaire et une autre plus contextuelle, au plus près des situations vécues par les usagers.

2.1. Protection de la vie privée et confidentialité

2.1.1. Utilisation massive des données et réglementations

Les pratiques de collecte et de traitement des données s'accroissent dans un contexte d'omniprésence de données constituant une ressource essentielle et précieuse auprès des entreprises et des gouvernements. Utilisées dans une visée de

2. Parmi nos références, notre webographie constitue une source importante compte tenu du développement des technologies vocales et de leur intégration récente dans les foyers. Afin de garantir une validité des données récoltées dans cette revue de l'état de l'art sur les interactions vocales, ainsi que les critères et les préconisations éthiques portés par les législateurs et les comités de réflexion, nous nous appuyons sur la triangulation des données visant une multiplication et une combinaison des matériaux récoltés. Cette triangulation nous permet notamment d'obtenir une connaissance plus complète et différenciée du même phénomène étudié (Berger *et al.*, 2010 ; Pinard *et al.*, 2004 ; Fortin, 1996).

profilage et de prédiction, ces données constituent le nerf de la guerre et jouent un rôle clé pour la compétitivité et la croissance dans les domaines du commerce, de la sécurité nationale, de la surveillance publique et de la recherche (Comest, 2017). Au risque de réduire les personnes concernées à de simples données instrumentées, l'utilisation massive de ces données à caractère personnel (photos, vidéos, voix, paramètre physiologiques, géolocalisation, etc.) peuvent constituer une forme d'intrusion, et ainsi entraver des droits et des libertés fondamentales telles que le respect de la vie privée (EDPS, 2015 ; Allistene, 2014). Afin de protéger les données personnelles et la vie privée des personnes, des réglementations et un cadre législatif ont été introduit depuis 1948, nous retrouvons tout d'abord l'article 12 de la Déclaration universelle des droits de l'homme (1948) stipulant : « *Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Toute personne a droit à la protection de la loi contre de telles immixtions ou de telles atteintes* » (Comest, 2017). Concernant la protection des données, la loi Informatique et Libertés de 1978 contient un ensemble de dispositions qui promeut une minimisation de la collecte et de la conservation des données personnelles comme une garantie nécessaire à la protection des personnes et de leurs libertés (CNIL, 2017-b). Dans le prolongement de cette loi, une initiative a été suivie tout récemment par la mise en place du règlement général sur la protection des données (RGPD) adopté par le Parlement européen le 14 avril 2016 et appliqué dans l'Union européenne depuis le 25 Mai 2018. Ceci afin d'accroître cette protection des données et responsabiliser les acteurs de ce traitement. Il instaure entre autres pour les personnes, le consentement explicite, des principes de protection dès la conception avec une sécurité par défaut, le droit à l'oubli en permettant la suppression de ses propres données, le droit à la portabilité des données et le droit à la limitation du traitement (Journal officiel de l'Union européenne, 2016). Par ailleurs, en cas de risques élevés pour les droits et les libertés des personnes concernées (exemples : situation de profilage par les données, surveillance systématique, usage innovant d'une technologie, décision automatique avec effet légal, etc.), ce règlement prévoit l'obligation de mener une analyse d'impact visant à décrire les opérations de traitements envisagés, une évaluation de cette nécessité des opérations, une évaluation des risques et les mesures envisagées pour faire face aux risques (Journal officiel de l'Union européenne, 2016 ; CNIL, 2017-b).

Toutefois, le rapport Villani (2018) souligne à juste titre un « angle mort du droit » qui n'est pas couvert par la législation actuelle. En ne traitant et en ne protégeant que les données à caractère personnel, l'exploitation des données collectives échappe ainsi à la portée de ces protections. D'où l'intérêt de développer des droits collectifs d'après Villani *et al.* (2018), afin d'adapter la protection des droits et des libertés à la logique technique de ces systèmes d'IA fonctionnant sur des quantités massives de données.

D'autres préconisations sont proposées en vue de protéger les données et la vie des personnes : mettre en place en amont du processus de collecte et de traitement des techniques d'anonymisation des données personnelles, ce principe peut ainsi être intégré « *dès la conception dans les nouveaux produits et services en masquant les données dans la mise en œuvre des algorithmes* » (EDPS, 2015). Un autre moyen

consisterait à octroyer un contrôle plus global des données auprès des personnes concernées en leur proposant des espaces de stockage sécurisés : « *un coffre-fort de données* » dont l'accès et l'usage est autorisé par les utilisateurs (op. cit.). Les systèmes informatiques et d'IA devraient également permettre la confidentialité des données, la sécurisation des transmissions et des stockages, le traçage des données, une transparence sur le traitement des données réalisé, ainsi qu'un droit d'opposition, d'accès et de rectification (Allistene, 2014 ; CNIL, 2017-b). Afin d'illustrer les initiatives encourageant la limitation de transmission et l'usage local des données, nous pouvons citer les travaux de la CNIL en collaboration avec la filière automobile. Dans une visée de protection des données personnelles des usagers de l'automobile connectée, ces acteurs ont élaboré un « *pack de conformité* » constituant une ligne directrice pour le marché européen des véhicules connectés. Ce pack intègre notamment un scénario impliquant une maîtrise totale des données par l'utilisateur : un stockage local au sein de réseaux de communication sous la maîtrise de l'utilisateur, une désactivation par défaut de la collecte en local, une possibilité d'accéder et de supprimer aisément l'historique des données (via un bouton accessible rapidement dans le véhicule et/ou le tableau de bord), informer l'utilisateur de la nature des données conservées et de leur traitement, proposer une collecte et un traitement des données en temps réel et sans stockage y compris en local, etc. (CNIL, 2017-b).

2.1.2. Assistants vocaux : renoncer à la confidentialité au profit de l'efficacité

L'utilisation des interfaces vocales, et plus précisément celui des assistants vocaux, invite à questionner l'usage éthique des données. En effet, la confidentialité et la sécurisation des données sont mises à mal par ces technologies qui doivent cumuler une quantité massive de données personnelles pour fonctionner efficacement (Davis, 2016). Dans un souci d'efficacité du dispositif et de contextualisation des services, les utilisateurs sont encouragés à renoncer à la confidentialité totale de leurs données dont on recommande un stockage permanent sur les serveurs afin d'assurer une mémoire des actions et des requêtes (Portet *et al.*, 2012 ; Davis, 2016 ; Metwally et Gregory, 2017). Certains utilisateurs évoquent une intrusion dans l'intimité liée à l'enregistrement continu des activités et des paroles. Ce sentiment d'intrusion ne peut être amoindri en cas de suppression des données sur l'historique, car l'ensemble du service serait mis à mal. Les utilisateurs éprouvent notamment un inconfort à diffuser des requêtes en présence d'autres personnes, ce qui rendrait la commande vocale difficilement acceptable dans un contexte collectif. Dans la continuité de ces problématiques sur la privauté des données, les utilisateurs regrettent également l'absence d'un bouton « *mute* » physique sur certains dispositifs : « *Je n'aime pas parce que j'ai la sensation d'être toujours écouté, et je dois rester prudente sur ce que je lui dis* » (Luria *et al.*, 2017). Cette question de la confidentialité est également soulevée en contexte d'usage partagé d'un assistant vocal au sein du foyer : ce qui pourrait occasionner l'intervention spontanée du dispositif pour proposer un produit ou un service dans un moment inadapté ou en présence d'autres membres ou de personnes extérieures.

Tous ces éléments discutés démontrent bien l'importance d'intégrer les problématiques de confidentialité dès la conception afin de renforcer l'acceptabilité de ces systèmes et faire en sorte que les utilisateurs puissent bénéficier de la totalité des services souhaités sans pour autant renoncer à la privauté, ni à la discrétion de leurs données.

Améliorer la sécurité et la confidentialité des données implique notamment une transparence du fonctionnement technique auprès des utilisateurs. Lorsque subsiste un manque d'intelligibilité sur ces questions, des défaillances peuvent survenir et contraindre l'utilisabilité et l'acceptabilité de ces dispositifs. Pour illustrer ce type de défaillance, nous citerons la logique de fonctionnement technique de l'assistant vocal Alexa. Cet appareil intègre des paramètres par défaut que bien souvent, les utilisateurs ne connaissent pas d'emblée ou n'en saisissent pas toutes les subtilités et les impacts possibles sur leurs activités. Parmi ces défauts de paramétrage, la sécurisation du compte personnel : le paramétrage par défaut n'inclut pas un code pin à 4 chiffres pour éviter des requêtes d'achats qui ne proviennent pas de l'utilisateur principal. Et Alexa autorise n'importe quelle personne proche du dispositif à commander des appareils et des services depuis le compte Amazon de l'utilisateur principal, et cela inclut les enfants, les voix de la radio ou de la télévision (Cobb, 2017). Ainsi, tant qu'un paramétrage des différentes voix n'est pas réalisé au sein du foyer, l'usage collectif de l'assistant vocal peut laisser entrevoir les requêtes et les actions des autres membres du foyer via l'historique des actions, les recommandations spontanées de l'assistant ou encore les notifications envoyées sur le compte de l'utilisateur principal.

Ce type d'usage nous invite ainsi à distinguer deux sortes de confidentialité à intégrer en conception : il s'agit de la *discrétion* – renvoyant au souhait de ne pas être entendu des autres personnes lors des interactions vocales (échelle locale) – et de la *privauté* – suggérant la protection et la sécurisation des données personnelles stockées par les dispositifs (échelle plus globale) –. Afin de pallier à ces contraintes, il demeure primordial d'intégrer dès la conception de ces dispositifs vocaux un bouton « *mute* » à part entière et d'assurer une confidentialité des données récoltées et stockées sur un serveur protégé. Privilégier une utilisation multi-utilisateurs (profils, droits d'accès) via la différenciation des voix permettra notamment d'assurer une meilleure discrétion des données. Il est donc important de faciliter au maximum ce paramétrage. Enfin, dans une visée de contrôle des données et dans l'objectif de rendre l'utilisateur acteur de l'utilisation de ses données, il est également nécessaire de favoriser l'exploitation et la modification des données personnelles et de rendre visible ces données recueillies et utilisées par le système.

2.2. Diversité culturelle et servicielle

2.2.1. Algorithmes de recommandation : services personnalisés ou bulle filtrante ?

Les réseaux sociaux, l'internet des objets, la diffusion du smartphone, l'open data et la dématérialisation des transactions et des services favorisent le développement d'une masse de données disponible dans une diversité de secteurs.

Ces mégadonnées soulèvent ainsi des enjeux de tri, de sélection et d'analyse auxquels les algorithmes sont en capacité de répondre afin de filtrer l'information, modéliser des phénomènes et profiler les individus. Le risque de ce type d'algorithme amènerait vers une sélection permanente de produits et de services provoquant un état d'isolement intellectuel en proposant aux utilisateurs ce qu'ils aiment et ce qu'ils pensent déjà, allant ainsi à l'encontre d'une diversité culturelle et d'un comportement ouvert et pluraliste. En introduisant le concept de « bulle filtrante », l'essayiste et militant internet Eli Pariser décrit ces dérives : « *Les filtres de personnalisation servent une sorte d'auto-propagande invisible : nous endoctriner avec nos propres idées, amplifier nos désirs pour des choses familières, et nous laissant inconscients des dangers cachés des territoires inconnus* » (Pariser, 2011). Les effets de ces algorithmes multipliés depuis l'émergence du Big Data empêcheraient les usagers de se diversifier, de s'ouvrir à d'autres univers et de se confronter à des positions contradictoires (Cardon, 2015).

Une façon de surmonter cet « enfermement algorithmique » pourrait consister à développer en parallèle d'autres algorithmes visant l'exposition à la diversité culturelle, de services et d'opinions pour répondre aux intérêts des utilisateurs, de manière à s'appuyer davantage sur les désirs et aspirations des usagers, plutôt que sur leur consommation quotidienne (CNIL, 2017-a).

2.2.2. Un dispositif faisant obstacle à la diversité de choix

Dans la continuité de ces problématiques, des utilisateurs évoquent un manque de diversité de choix dans les propositions spontanées et contextualisées de certains assistants vocaux (Davis, 2016 ; Cobb, 2017). Davis (2016) évoque plusieurs paramètres inhérents au dispositif vocal Alexa intégrant des recommandations (dont la logique de fonctionnement reste opaque pour les utilisateurs) qui sont proposées aux utilisateurs sans choix possibles : l'assistant vocal propose des achats à partir de l'historique de l'utilisateur, du choix d'Amazon (meilleures notations, meilleurs prix, etc.) et des items éligibles lors de la commande « *prime* ». Ce dispositif se retrouve donc dans une position privilégiée et unique pour dicter des choix de produits et de médias (*op. cit.*). Ce fonctionnement s'appuie sur l'usage d'algorithmes de recommandation pouvant non seulement empêcher l'utilisateur de disposer d'une liberté de choix dont il disposerait en délaissant les interactions vocales, mais ces algorithmes peuvent notamment déployer une bulle filtrante autour des usagers. Ce qui les amènerait à consommer des produits ou des services, à se divertir et à s'informer le plus souvent à partir des mêmes sources sans s'ouvrir à d'autres univers ou d'autres opinions (Pariser, 2011). S'appuyant sur un traitement de masse et des usages collectifs, ces algorithmes ne sont pas non plus assurés de fournir des recommandations pertinentes pour chaque utilisateur. Cette masse de données hétérogène et non triée éradique non seulement les significations des événements récoltés, mais ils désobjectivent notamment les personnes concernées tout en considérant les sujets comme des réseaux de données recombinaisons. Cette façon de gouverner par les algorithmes éradique ainsi tout ce qui fait l'essence de l'humain : ses rêves, ses espoirs, les parts d'incertitudes, les critiques, les projections, le devenir de chacun, les réflexions et le sens que l'on donnerait aux événements (Rouvroy et Berns, 2013). Ce traitement en masse empêche surtout un traitement

contextuel des données susceptible de répondre de façon pertinente aux requêtes implicites, aux langages spécifiques et aux besoins présents des utilisateurs non exprimés.

Par ailleurs, la diversité servicielle du dispositif vocal peut notamment être mise à mal lorsqu'il existe un manque d'intelligibilité sur les potentialités des services vocaux. Cela nuit à l'expérience utilisateur lorsque les usagers ne connaissent pas l'étendue, ni les limites fonctionnelles de leurs appareils. Ils évoquent ainsi une déception et une mauvaise expérience lorsqu'ils découvrent les limites de certains *skills* d'assistants vocaux ne supportant pas par exemple les requêtes complexes impliquant un ensemble de facteurs : « *En lançant l'application « tremblement de terre », on ne peut pas demander le plus gros tremblement de terre qui a eu lieu la semaine passée et l'application envoie une information erronée* » (Rasmus, 2016). De plus, une interaction exclusivement vocale combinée à une méconnaissance des actions possibles peut limiter la capacité des utilisateurs à explorer le système, ou encore nuire aux capacités de réflexion dans l'action retrouvées habituellement dans les interactions physiques (Klemmer *et al.*, 2006) : « *Pour utiliser l'interface vocale, on doit s'arrêter, et réfléchir à ce qu'on doit dire* » (Luria *et al.*, 2017).

Cette problématique met en lumière l'importance de l'intelligibilité et de la clarté du fonctionnement algorithmique à renforcer auprès des utilisateurs. Cette meilleure connaissance des potentialités et limites permettrait ainsi aux usagers qui le souhaitent de garder la main et de modifier si besoin la provenance des sources utilisées sur les recherches du web et sur les services de recommandation. Des préconisations pourraient également être orientées vers la proposition de « *skills* » et de fonctionnalités encore inconnues par l'utilisateur susceptibles de lui être utiles.

2.3. Justice, équité et diversité des profils utilisateurs

2.3.1. Biais algorithmiques et risques de discrimination

L'actualité regorge d'évènements liés aux biais algorithmiques provoquant de nombreux enjeux éthiques : nous retrouvons par exemple l'expérience du bot conversationnel « Tay » de Microsoft rapidement retiré des réseaux sociaux en Mars 2016. Ce robot conversationnel incarnait une adolescente sur Twitter capable d'accroître son intelligence et d'apprendre de ses interactions avec les internautes. Toutefois, le programme a dû être suspendu suite aux propos racistes, sexistes et négationnistes qu'elle tenait après quelques heures d'échange et d'entraînement au contact des internautes. Ce sont donc les données fournies aux systèmes qui ont soulevé des problématiques en reproduisant les discriminations préexistantes (Tual, 2017 ; CNIL, 2017-a). Nous pouvons également citer une étude controversée menée par des chercheurs américains de Stanford consistant à « *entraîner une IA à reconnaître des homosexuels depuis de simples portraits d'utilisateurs d'applications de rencontre* » (Guillaud, 2017). Pour réduire ces discriminations, il importe donc d'examiner les questions de justice et d'équité au sein même des algorithmes pour ne pas reproduire les préjugés et les stéréotypes de toutes sortes (Comest, 2017). Le rapport de la Commission nationale Informatique et Libertés

(CNIL, 2017-a) sur les enjeux éthiques des algorithmes et de l'IA nous rappelle que tout algorithme demeure en partie biaisé car il reste « *le reflet – à travers son paramétrage et ses critères de fonctionnement, ou à travers les données d'apprentissage qui lui ont été fournies – d'un système de valeurs et de choix de société (...)* Celui qui entraîne un algorithme y insère d'une certaine façon sa propre vision du monde, ses valeurs ou, à tout le moins, des valeurs présentes plus ou moins directement dans les données tirées du passé ». Souvent invisibles, ces biais sont d'autant plus prégnants avec les algorithmes d'apprentissage automatique qui reproduisent toutes les discriminations connues dans nos sociétés lorsque les données ne sont pas corrigées (Rapport parlementaire, 2017).

Afin d'éviter la reproduction ou le renforcement de biais et de discriminations diverses, des propositions ont été émises par la CNIL (2017-a) : il s'agirait d'une part, d'assurer la qualité des données entrantes dans les systèmes algorithmiques et d'intelligence artificielle (éviter les données erronées, périmées, manquantes ; les données véhiculant ou reproduisant des biais préexistants ; prémunir les systèmes d'actes de malveillance volontaire). Pour y parvenir, le paramétrage des algorithmes demeure essentiel en prêtant attention aux critères à partir desquels ils fonctionnent, sélectionnent et recommandent. D'autre part, la CNIL propose d'intégrer un principe d'intelligibilité ou d'explicabilité des algorithmes visant l'accessibilité pour tous et la compréhension générale du fonctionnement de l'algorithme : « *que l'on comprenne ce qui rentre et ce qui sort de l'algorithme ainsi que son objectif. C'est cela qui doit être transparent* » (op. cit.).

2.3.2. Formes d'adressages et problématiques d'inclusion et d'exclusion des catégories d'utilisateurs

Certaines problématiques de discrimination et d'équité ont été évoquées par les utilisateurs concernant l'inadéquation des formes d'interaction pour s'adresser aux assistants vocaux. En effet, certains utilisateurs apprécient que certains assistants vocaux privilégient des formes d'adressage et d'interaction intégrant une forme de courtoisie et de respect des genres. D'une part, l'assistant vocal Alexa propose des formes d'interaction sur un ton injonctif dirigées vers une figure féminine (« *Alexa, éteins les lumières* » ; « *Alexa, programme mon réveil pour 8h* »). D'autre part, l'assistant Google Home invite davantage à converser, tout en proposant une interaction plus respectueuse (« *Ok Google, lance l'actualité* » ; « *Hey Google, programme mon réveil pour 8h* »). Une utilisatrice attire ainsi l'attention sur l'impact émotionnel des technologies sur l'humain et sur le modèle et les valeurs sociales véhiculées par certains assistants vocaux : « *La différence d'expérience vécue ici est énorme. Lorsque j'utilise Alexa après la Google Home, je me sens autoritaire (...)* Ce n'est pas isolé de la façon dont les hommes considèrent les femmes, mais cela s'étend aussi sur la manière dont les femmes se considèrent entre elles. Les moindres détails d'expérience utilisateur ont de l'importance (...) Je souhaite remercier Google. Merci d'avoir porté une attention à ces détails, et aux utilisatrices » (Paolino, 2017).

Afin de garantir une accessibilité auprès d'une variabilité d'utilisateurs, l'équité représente un point d'ancrage pouvant être intégré dès la conception. Dans cette

visée, la modalité vocale constitue une forme d'interaction utilisable par une grande variété d'utilisateurs et d'individus, dont les personnes en situation de handicap. La commande et la reconnaissance vocale intègrent potentiellement une adaptation inter-individuelle consistant à aménager des produits, des services ou des équipements susceptibles d'être utilisés par toute personne, peu importe le sexe, l'âge, la situation ou le handicap. L'interaction devient ainsi utilisable et mobilisable par une diversité de publics telles que les enfants, les illettrés, les personnes âgées, les malvoyants, ou les publics ayant des difficultés à interpréter certains graphiques ou certaines données complexes (Portet *et al.*, 2012 ; Munteanu et Penn, 2016 ; DeRenzi *et al.*, 2017). Concernant les enfants, ces derniers n'ont pas la dextérité ni une taille de main suffisante pour utiliser efficacement les claviers et les pointeurs. Et tous les environnements ne sont pas tous disposés à proposer des commandes via une manipulation directe. Ces contraintes peuvent être surmontées et les technologies vocales intégrées dans les usages à condition que les facteurs humains soient pris en compte lors des phases de conception (Gardner-Bonneau, 1999 ; Harris, 2005).

Néanmoins, il convient de bien distinguer ce potentiel associé à cette interaction a priori plus naturelle et spontanée pour l'utilisateur, à la réalité dans son utilisation effective. En effet, de nombreux dysfonctionnements ont été soulevés par diverses catégories d'utilisateurs. En premier lieu, des difficultés de reconnaissance vocale subsistent avec les voix des enfants, des personnes âgées, des personnes atteintes d'une maladie ou ayant des accents (Portet *et al.*, 2012 ; Soronen *et al.*, 2008). D'autres difficultés pointent une absence de fonctionnement sur une logique de parole libre : les biais d'interprétation sémantiques, les mauvaises reconnaissances et l'usage spécifique de mots-clés pour certaines *skills* augmentent le taux d'erreurs et amènent les utilisateurs à modifier et à adapter leur vocabulaire. Cela implique ainsi une phase d'apprentissage sur le long terme afin de personnaliser l'assistant vocal et augmenter son taux de réponses pertinentes. Cette particularité de fonctionnement s'adresse donc à des utilisateurs disposant des capacités et des connaissances antérieures suffisantes pour comprendre ces usages et s'y adapter.

Garantir une forme d'équité dans l'expérience d'usage nécessite d'identifier les variétés de profils utilisateurs et d'évaluer l'impact de ces systèmes d'IA sur leurs activités tout en pointant les nombreux freins à l'acceptabilité de ces dispositifs. Des recherches sur les seniors démontrent que ces publics craignent de développer une passivité et une oisiveté dans les tâches quotidiennes qu'ils préfèrent prendre le temps de réaliser manuellement (Soronon *et al.*, 2008 ; Portet *et al.*, 2012). Il est donc impératif de privilégier des tentatives d'aide à la personne qui ne réduisent ni le sens du contrôle, ni leur propre estime (Rodin, 1986). Les travaux de Portet *et al.* (2012) sur les systèmes vocaux en situation domestique soulignent bien l'importance de préserver auprès des personnes âgées ce sentiment de contrôle et leur capacité à agir en toute autonomie : « *J'aime bien agir plutôt que parler (...) J'aime bien fermer mes volets, etc. (...) Moi en ce moment je préfère faire les choses parce que sinon c'est glisser vers l'inactivité. Faudrait vraiment que je puisse plus le faire, parce que sinon on ne fait plus rien, on se couche et puis voilà* » (op. cit.). Quant à l'utilisation des assistants vocaux par des enfants, des parents évoquent plusieurs contraintes telles que le risque de généraliser une forme d'impolitesse lorsque ces

enfants entendent les requêtes exécutées sans les mots « s'il te plaît » et « merci ». Ou encore, conséquence d'une intégration totale de cette forme d'interaction, les parents observent avec surprise une tendance auprès de jeunes enfants (2 ans) à mobiliser l'interaction vocale avec une variété d'objets non connectés dans le foyer incluant par exemple la télévision ou l'horloge de la cuisine (Heritage, 2017).

Pour répondre à ces enjeux de justice, d'équité et d'inclusion d'une diversité d'utilisateurs, plusieurs préconisations sont possibles : proposer un choix d'adressage vis-à-vis de l'assistant vocal (mot-clé d'activation, forme de requête, types de voix, etc.). Il s'agira notamment de renforcer le sentiment de contrôle et la capacité d'agir auprès des seniors dans l'expérience d'usage du vocal : cela nécessite un travail d'investigation en situation réelle afin d'identifier les besoins et les cas d'usages dans lesquels le vocal ne porte pas atteinte aux valeurs et aux principes de ces publics, tout en venant soutenir ou alléger certaines activités complexes ou pénibles à mettre en œuvre. Pour les enfants, des propositions peuvent être orientées vers une intégration d'une courtoisie dans l'interaction comme condition sine qua non pour l'acceptation des requêtes. Il sera également déterminant d'intégrer une logique de parole libre et d'améliorer les capacités de détection et de reconnaissance des voix spécifiques (enfants, seniors, accents, etc.). Enfin, nous recommandons de privilégier une proactivité du système dans l'accompagnement des utilisateurs pour la prise en main de l'assistant et des usages sur le long terme (guidage, propositions de nouveaux *skills*, didacticiel, etc.).

2.4. Autonomie du système et reprise en main par les utilisateurs

2.4.1. Combiner une autonomie à une prévisibilité du système

L'autonomie d'un système d'IA ou d'un robot représente sa capacité à fonctionner pendant une durée prolongée, et indépendamment d'un humain dans des environnements complexes et dynamiques à l'aide des nombreux capteurs leur permettant d'interpréter toutes sortes d'informations (Allistene, 2014 ; Bekey, 2012). Parmi cette capacité, nous retrouvons deux catégories d'autonomie : la première opérationnelle, est orientée vers la perception et l'action du système ; la seconde décisionnelle, concerne la capacité de raisonnement sur la perception et les actions permettant au système de réaliser des choix (Allistene, 2014). Cette capacité d'autonomie contribue à renforcer l'imprévisibilité des systèmes d'IA qui « *ne sont pas simplement programmés pour exécuter des tâches spécifiques mais pour apprendre et continuer à se développer en interaction avec leur environnement* » (Comest, 2017). Bien que cela facilite l'exécution de tâches complexes, un haut niveau d'autonomie soulève des enjeux éthiques de prévisibilité et de responsabilité juridique pouvant impacter les utilisateurs.

En cas de situation d'urgence où l'échange avec l'humain est impossible ou en cas de situation critique dans laquelle l'homme n'a pas de contrôle sur le système et que le contexte nécessite une prise de décision imminente, quel type de comportement faudrait-il prétablir ? Quelle règle éthique intégrer en amont au sein

du système pour l'amener à prendre la bonne décision selon le contexte³ ? Ou encore, quand et comment faut-il transférer cette capacité d'agir et de reprise en main à l'utilisateur concerné ?

Non seulement, il est possible d'intégrer des règles de conduite morale pour encadrer la conduite du système, mais il est également essentiel d'intégrer une fonction de reprise en main par l'utilisateur pour qu'il soit en mesure d'intervenir dans n'importe quelle circonstance tout en étant informé de la situation en cours. La fonction de débrayage des fonctions autonomes du système représente également un principe à privilégier et à rendre accessible aux utilisateurs (ALLISTENE, 2014). Afin d'anticiper les situations néfastes ou critiques, il importe de proposer aux utilisateurs une documentation claire qui informe de la logique de fonctionnement (système de discrimination de situation, d'interprétation, mobilisation de codes moraux, etc.), des capacités et des limites du système d'IA. En vue de faciliter la reprise en main par les utilisateurs, il demeure nécessaire de les informer sur les décisions prises par le système afin de favoriser une continuité dans la compréhension de la situation en cours (op. cit.). Pour empêcher une forme d'autonomie incontrôlable, il est important de favoriser une prévisibilité du système humain-robot en délimitant et en rendant transparent le périmètre d'action et les défaillances possibles du système pour l'utilisateur lui permettant de construire une compréhension de la situation et des décisions prises par le système (ALLISTENE, 2014 ; CNIL, 2017-a ; Villani *et al.*, 2018).

2.4.2. Renforcer l'intelligibilité, l'explicabilité et le sentiment de contrôle dans l'usage du vocal

Parmi les défaillances du système vocal, nous avons précédemment évoqué le manque d'intelligibilité du système, notamment sur des questions sécuritaires importantes (exemple : un paramétrage par défaut sans code pin pour protéger le compte des requêtes extérieures et des commandes de produits sur le compte Amazon de l'utilisateur principal) (Cobb, 2017). Concevoir des interactions qui favorisent une intelligibilité et un sentiment de contrôle auprès des utilisateurs peut être facilité en s'appuyant sur la multimodalité des formes d'interaction afin que ces modalités se soutiennent les unes aux autres selon les contextes et les préférences des usagers. À ce titre, des utilisateurs des technologies vocales soulèvent des difficultés telles que l'absence d'écran visuel pour soutenir le vocal, et qui potentiellement impacte le sentiment de contrôle et amoindrit la possibilité de reprendre la main sur l'appareil en cas de besoin (Luria *et al.*, 2017 ; DeRenzi *et al.*, 2017 ; Metwally et Gregory, 2017 ; Whinton, 2016). Une étude réalisée par Luria *et al.* (2017) portant sur des dispositifs domestiques intégrant une diversité de

3. Ces questions éthiques liées à l'autonomie sont souvent illustrées par les exemples du tramway dévalant sans freins sur une pente ou de la conduite autonome afin de soulever la complexité de règles à appliquer en situation de dilemme éthique : en cas de risque d'accidents, le tramway devrait-il s'engager sur une voie l'amenant à tuer une personne ou sur une seconde voie dans laquelle plusieurs personnes risquent d'être tuées ? Le système d'IA devrait-il risquer de percuter un piéton ou mettre en péril un passager dans l'habitacle ? (Allistene, 2014 ; Comest, 2014 ; CNIL, 2017-a ; Rapport parlementaire, 2017).

formes d'interaction (un haut-parleur à commande vocale ; un écran tactile mural, une application mobile ; un robot social) a justement démontré l'importance du canal visuel auprès des utilisateurs pour disposer d'informations sur la réussite ou non des commandes réalisées et pour obtenir des informations sur la situation en cours et le contexte de la machine : « *Au final, je ne sais rien avec le vocal. C'était si facile de réaliser les commandes, mais finalement vous ne saviez pas ce qui n'allait pas* ». Ces aspects favorisent ainsi une impression de manque de contrôle auprès des utilisateurs. Cette perte de contrôle représente un véritable frein à l'acceptabilité du vocal cité par certains utilisateurs. Ils évoquent une forme d'inconfort à perdre l'initiative et le contrôle de certaines fonctions ou usages d'objets domestiques : « *En utilisant le vocal, ça donne l'impression que c'est vous qui décidez mais quelqu'un d'autre décide aussi pour vous. Sur cet écran tactile, vous sentez que vous avez le contrôle à 100%, mais ici avec le vocal, c'est comme si vous attendiez qu'ils acceptent ce que vous demandez. L'inconvénient c'est qu'on ne se sent pas à 100% en situation de contrôle comme les autres interfaces. C'est comme si on sollicitait une chose à la maison et qu'on doit attendre que ce soit approuvé* » (Luria *et al.*, 2017).

Ce critère d'explicabilité à intégrer dès la conception conditionne l'utilisabilité, l'acceptabilité et le sentiment de confiance vis-à-vis du dispositif. De nombreux travaux ont démontré que la quantité et la pertinence des explications sur la logique de fonctionnement du système, sur son mode de « *raisonnement* » et la source des problèmes renforcent tout d'abord la confiance dans le système. Par ailleurs, ces connaissances permettent également aux utilisateurs de mieux comprendre le système en cas d'erreurs, reprendre la main afin de le déboguer si besoin et améliorer le fonctionnement du système (Kulesza *et al.* 2013 ; Bussone *et al.*, 2015). Favoriser une compréhension des comportements observables du système, au cours de l'action et non a posteriori (Stumpf, 2015), renforcerait ainsi la prise en main et l'appropriation du dispositif pour permettre aux utilisateurs d'adapter les moyens d'actions selon la situation (en cas d'urgence ou non).

A contrario, nous retrouvons un cas d'usage où le sentiment de contrôle du dispositif peut faire émerger plus clairement des divergences sur les valeurs et les principes domestiques portées par les membres du foyer. Heritage (2017) nous relate ainsi une expérience d'usage collectif dans laquelle le dispositif Alexa est utilisé par la mère de famille pour réchauffer les pièces du foyer via le contrôle du thermostat. Et cela, de manière discrète en soirée, et à l'insu de son mari qui avait programmé en amont une température plus froide. Dans ce contexte, l'appareil devient un instrument de contrôle domestique permettant de prendre l'ascendant malgré un désaccord de principes au sein du couple : « *Elle préfère une maison accueillante et chauffée tandis que je préfère un foyer plus frais dans lequel tout le monde porte un tas de pulls parce que le chauffage est coûteux et que je paye les factures* » (*op. cit.*). En étant exacerbé, ce sentiment de contrôle a provoqué un partage de « *pouvoir d'action* » facilitant le pilotage d'objets connectés au sein du foyer, tout en faisant émerger des problématiques de répartition de pouvoir et de négociation au sein du couple.

Ces éléments abordés soulèvent de nouveau l'importance d'une intelligibilité et des informations à transmettre auprès des usagers sur le fonctionnement des systèmes d'IA, en particulier pour les points critiques du système liés à la sécurité et à la confidentialité des données. Et comme évoqué précédemment, favoriser une multimodalité des interactions constitue un moyen de diversifier des façons de reprendre la main sur le système tel que l'ajout du canal visuel, l'intégration d'un bouton « *mute* » physique pour débrayer plus facilement et notamment faciliter les formes d'automatisation à l'aide de scénarios créés par les utilisateurs.

2.5. Biomimétisme et liens affectifs

2.5.1. Une IA affective : des risques d'anthropomorphisation à évaluer

Les futurs systèmes d'intelligence artificielle intégreront toute une palette d'expression d'émotions (sentiments, de plaisirs, de peurs, valeurs morales, etc.) combinée à des techniques de perception des émotions qui leur permettront de s'adapter à l'état d'esprit des utilisateurs et de raisonner en tenant compte de ces informations relatives aux émotions (Rapport parlementaire, 2017 ; CNIL, 2017-a ; Allistene, 2014). Développer ce champ de la robotique affective viserait ainsi à renforcer l'acceptabilité et le plaisir à interagir avec un système d'IA qui s'adapte, nous comprend et se synchronise à l'homme : « *Engager l'utilisateur dans une interaction avec un robot doté de capacités affectives, faire qu'il prenne du plaisir à interagir avec lui, faire qu'il ait confiance en lui, sont des buts de la robotique d'assistance* » (Allistene, 2014). Toutefois, donner l'impression qu'un système d'IA possède des émotions risquerait d'amener les utilisateurs à attribuer aux robots ou aux systèmes d'IA une subjectivité et une sensibilité (Turckle, 2015), tout en provoquant la création d'un lien affectif avec l'IA (Tisseron, 2015 ; CNIL, 2017-a).

Ces enjeux, dont on peut aisément questionner l'acceptabilité morale nécessitent de rester vigilants sur les effets occasionnés par ces formes d'anthropomorphisation, de questionner l'utilité de ces ressemblances au vivant, notamment auprès de publics plus fragiles tels que les enfants et les personnes âgées qui ont tendance à s'attacher plus facilement aux robots (Epley *et al.*, 2007 ; Allistene, 2014). Il demeure également important de communiquer de façon transparente sur les répercussions possible auprès des utilisateurs et d'évaluer l'impact de ces formes d'interaction sur leur affectivité (Allistene, 2014).

2.5.2. Technologies vocales : quand le malaise et l'inquiétude s'invitent auprès des utilisateurs...

Les utilisateurs mentionnent une forme de déstabilisation récurrente avec les technologies vocales : il s'agit d'un malaise et d'une crainte provoqués par une sensation de présence liée à cette parole qui s'apparente à une présence humaine : « *Je n'aime pas le fait de me parler à moi-même, ça me donne une impression étrange et très bizarre... Je ne sais pas, c'est bizarre d'avoir une voix qui vous répond. C'est un peu effrayant* » ; « *Je ne peux pas vraiment expliquer pourquoi, mais je ne me sens pas à l'aise avec. Parce que c'est une machine et qu'il peut me répondre. Je me*

sens juste inconfortable » ; « *Il pourrait devenir une sorte d'amis au bout d'un certain temps. On pourrait se sentir très à l'aise avec, comme si on était plus seul ? Peut-être que c'est un peu effrayant* » (Luria et al., 2017). Ces technologies biomimétiques révèlent ainsi le risque de certaines technologies et intelligence artificielle à brouiller les frontières entre un être vivant et un artefact (Allistene, 2014).

Ce malaise provoqué par cette sensation de présence peut être d'autant plus marqué lorsque le dispositif vocal se met à fonctionner indépendamment des commandes ou du libre choix des utilisateurs. Pour illustrer cet aspect, nous pouvons citer un problème technique cité récemment dans l'actualité en Mars 2018 : un ensemble de haut-parleurs Alexa dans les foyers se mettaient à rire spontanément sans aucune raison apparente, et sans aucun lien avec une commande vocale, et cela à n'importe quel moment de la journée. Les utilisateurs ont proposé une diversité de qualificatifs pour décrire ce rire pouvant être « sardonique » ; « effrayant » ; « inquiétant » ; « glaçant » ; ou encore « démoniaque intempestif ». Voici quelques verbatims des utilisateurs ayant vécu cette défaillance technique : « *J'étais dans mon lit et sur le point de m'endormir quand Alexa a lancé un rire fort et terrifiant. Il y a de grandes chances que je sois assassiné cette nuit* » (Charnay, 2018) ; « *Donc, Alexa a décidé de rire aléatoirement pendant que j'étais dans la cuisine. Je pensais qu'un gamin riait derrière moi* » ; « *On était en train de parler de choses assez confidentielles et Alexa s'est mise à rire. Quelqu'un d'autre a déjà eu ça ? On n'avait pas l'impression de l'avoir déclenchée accidentellement, elle s'est juste mise à rire comme ça. C'était très effrayant* » (Hillard, 2018). Vécue et jugée comme une expérience désagréable, voire effrayante par les usagers, ce bug technique renforce également ce sentiment de malaise par les enjeux d'anthropomorphisation des technologies que cela soulève.

Une façon de surmonter ces inquiétudes liées aux risques de l'anthropomorphisation pourrait consister en un choix de paramétrage entre un « *assistant fonctionnel* » ou un « *assistant de compagnonnage* ». Ce dernier choix proposerait une variété de critères et d'attributs que l'utilisateur pourrait ajouter, ajuster ou enlever selon ses préférences : variation de voix ou de timbre, intégration de marques de politesse et/ou d'humour, sélectionner un type de personnalité pour l'IA, expression et reconnaissance des émotions, etc.

2.6. Quels enseignements tirer de cette mise en perspective ?

Cette mise en perspective nous a permis de mettre en lumière des réalités et des enjeux parfois similaires ou divergents concernant l'éthique des systèmes d'IA en nous appuyant sur les visions des législateurs et celles des utilisateurs. Une triple approche intégrant la description des constats, des retours d'expériences (REX) et des préconisations vise à obtenir une vision systémique de l'éthique pour en délimiter ses contours, tout en l'illustrant d'enjeux concrets avec des personnes éprouvant et expérimentant les limites ou les bénéfices d'une éthique intégrée à des systèmes d'IA. À partir du tableau 1 récapitulant les principaux éléments, nous proposons d'extraire des pistes de réflexion et d'en tirer des enseignements.

Tableau 1. Synthèse des visions réglementaires et contextuelles sur l'éthique de l'IA depuis les perspectives des législateurs et des utilisateurs

CRITÈRES ÉTHIQUES		LÉGISLATEURS	UTILISATEURS (Technologies vocales)
1. Protection de la vie privée et confidentialité	Constats et REX	<ul style="list-style-type: none"> - Risques des données instrumentées : intrusion, entraver les droits et libertés - Cadre réglementaire existant (Déclaration des droits de l'homme & Loi informatique et libertés + RGPD) 	<ul style="list-style-type: none"> - Efficience du système au détriment de la confidentialité - Sentiment continu d'intrusion dans l'intimité - Crainte que la discrétion des données ne soit pas préservée
	Préconisations	<ul style="list-style-type: none"> - Technique d'anonymisation - Droit à l'oubli - Droit à la limitation du traitement - Consentement explicite - Portabilité des données - Coffre-fort de données - Analyse d'impact éthique - Pack de conformité pour les véhicules autonomes - Créer des droits collectifs pour la protection des données 	<ul style="list-style-type: none"> - Intégration d'un bouton mute - Stockage sur un serveur sécurisé et stockage en local - Différenciation des voix - Paramètre de contrôle pour l'utilisateur principal - Transparence sur le fonctionnement du système - Paramétrage de sécurité par défaut sur le dispositif
2. Diversité culturelle et servicielle	Constats et REX	<ul style="list-style-type: none"> - Risques des algorithmes de recommandation : profilage réducteur des individus, manque de diversité culturelle et servicielle, « bulle filtrante » 	<ul style="list-style-type: none"> - Manque de choix dans les propositions de services et de produits des assistants vocaux (type Alexa) - Hégémonie du système dictant des choix de produits/médias (contrainte de l'interaction vocale exclusive) - Recommandations pas pertinentes pour tous - Méconnaissance des skills, des limites et potentialités du système (limite du vocal)
	Préconisations	<ul style="list-style-type: none"> - Encourager des algorithmes qui exposent à la diversité - S'appuyer sur les désirs/aspirations VS consommations routinières 	<ul style="list-style-type: none"> - Favoriser une multimodalité des interactions pour l'exploration et la diversité de choix - Intelligibilité et explicabilité du comportement du système - Faciliter et encourager la modification des sources utilisées pour les recherches d'information et les services de recommandation (choix entre des algorithmes « ouverts » / « fermés ») - Proposition de nouvelles skills pour les usagers

3. Justice et équité	Constats et REX	<ul style="list-style-type: none"> - Risques de discriminations algorithmiques : problématiques d'inclusion et d'exclusion de profils, racisme, sexisme, etc. - Paramétrage et données d'apprentissages biaisés 	<ul style="list-style-type: none"> - Formes d'adressages et d'interactions verbales pouvant induire des stéréotypes liés au sexe (ton d'injonction, voix féminine, etc.) - Interaction vocale invitant à une expérience adaptée à une diversité de profils - Problème de reconnaissance (seniors, enfants, malades, accents) - Impact différents selon les profils : <i>seniors</i> (sentiment de passivité, manque d'estime et d'autonomie), <i>enfants</i> (risque de généraliser l'impolitesse, mobiliser l'interaction vocale sur d'autres objets non adéquats) - Pas de parole libre - Phase d'apprentissage
	Préconisations	<ul style="list-style-type: none"> - Vérifier la qualité des données entrantes lors de l'apprentissage - Vigilance sur les critères des algorithmes mobilisés - Rendre les algorithmes intelligibles et explicables 	<ul style="list-style-type: none"> - Proposer un choix d'adressage auprès de l'assistant (mot-clé d'activation, forme des requêtes, types de voix, etc.) - Améliorer les détections et reconnaissances de voix - Particularités d'usages et impacts différents selon les profils : <i>seniors</i> (renforcer le contrôle et capacité d'agir), <i>enfants</i> (renforcer l'adressage courtois et prêter vigilance aux impacts du vocal sur le développement) - Intégrer une logique de parole libre - Proactivité pour accompagner la prise en main du système (guidage, skills, didacticiel)
4. Autonomie et reprise en main	Constats et REX	<ul style="list-style-type: none"> - Imprévisibilité des systèmes d'IA autonomes - Problème de responsabilité en cas de défaillances 	<ul style="list-style-type: none"> - Manque de transparence sur le fonctionnement, les potentialités et limites des assistants vocaux - Intégration de paramétrages de sécurité par défaut non intelligible pour l'utilisateur et potentiellement dangereux - Sentiment de contrôle amoindri par l'interaction exclusive du vocal - Cas d'excès de contrôle : contrainte dans les négociations de pilotage du thermostat, survenue de conflits de valeurs
	Préconisations	<ul style="list-style-type: none"> - Renforcer la transparence du système pour l'utilisateur - Intégrer des règles de conduite morale au système dès la conception - Multiplier les possibilités de reprise en main et de débrayage du système 	<ul style="list-style-type: none"> - Intelligibilité et information à renforcer auprès des usagers sur les points critiques du système (ex : confidentialité, sécurité) - Favoriser une multimodalité des interactions pour diversifier les possibilités de reprise en main (ajout du canal visuel, intégration d'un bouton mute physique pour débrayer, faciliter les automatisations avec des scénarios) - Intégrer de l'explicabilité pour : comprendre la logique du système, les erreurs, permettre de déboguer, reprendre la main

5. Biomimétisme et liens affectifs	Constats et REX	- Risque d'attribuer des émotions aux systèmes d'IA et de créer un lien affectif (publics fragiles)	- Malaise provoqué par une sensation de présence liée aux voix des assistants vocaux - Défaillance du système qui intervient spontanément de manière effrayante
	Préconisations	- Informer sur les risques - Questionner l'utilité de l'anthropomorphisation	- Paramétrer les formules d'activation, apprendre au système un vocabulaire propre - Proposer aux utilisateurs d'ajouter, ajuster ou supprimer des critères et des attributs liés à la voix (timbre, variation, type de voix, personnalité, marque politesse, etc.) => le rendre – ou + humanisant selon les préférences des usagers

Au-delà du cadre réglementaire, les retours d'expériences démontrent bien que les usagers utilisant les assistants vocaux réalisent des compromis sur la confidentialité des données afin de bénéficier d'un système vocal plus efficace (mémorisation de toutes les requêtes entraînant une contextualisation des services et une meilleure pertinence de réponses aux requêtes). Ces éléments semblent indiquer que l'acceptabilité de tels systèmes se rapporte à une question d'équilibre et de balance entre les contraintes rencontrées (liés aux valeurs, aux représentations et à l'activité des utilisateurs) et les intérêts dont bénéficient les usagers au profit des conflits éthiques auxquels ils peuvent être confrontés. Pour d'autres utilisateurs, la confidentialité représente une condition *sine qua non* d'une première intégration dans le foyer et d'une utilisation effective. Par ailleurs, les éléments discutés sur les enjeux de confidentialité, de diversité culturelle et servicielle, et d'autonomie pointent l'importance de proposer une plus grande transparence sur l'usage des données (*Que fait-on de mes données ?*) et davantage de contrôle sur les données sortantes (*Comment puis-je sélectionner et filtrer mes données sortantes ?*). Une autre forme de transparence est aussi évoquée, il s'agit de l'intelligibilité du système et l'explicabilité de son fonctionnement. Ces problématiques d'intelligibilité contribueraient à rendre les utilisateurs acteurs de leurs choix, des services utilisés et des paramétrages souhaités afin de renforcer une prévisibilité du système, ainsi qu'un sentiment de contrôle et de sécurité.

Étudiées en contexte d'utilisation réelle, les problématiques de justice et d'équité révèlent des enjeux non prévisibles par les concepteurs, tout en affectant l'activité des utilisateurs ainsi que l'acceptabilité des systèmes. Pour ces raisons, surmonter les discriminations, les stéréotypes et les préjugés divers pouvant être occasionnés par l'IA implique d'étudier l'activité et les expériences d'usage selon une variété de profils susceptibles de vivre les interactions avec l'IA de façon différenciée. Enfin, les enjeux du biomimétisme impliquent des dimensions émotionnelles et imaginaires imbriquées chez les utilisateurs dans l'expérience d'utilisation. Les principales difficultés relevées par les utilisateurs concernaient les interventions spontanées de l'IA qui n'étaient ni souhaitées ni prévues en situation, ce qui a pu entraîner une profusion de sentiments négatifs, de méfiance et d'inquiétudes vis-à-vis du système.

En perdant cette dimension du contrôle, cela entraîne une perte de confiance pouvant faire obstacle à l'acceptabilité de ce système.

3. Discussion et perspectives

Ce numéro spécial sur l'éthique et l'intelligence artificielle a représenté une opportunité riche et intéressante pour questionner des problématiques actuelles articulées à l'usage des algorithmes et de l'intelligence artificielle. Ces enjeux éthiques soulèvent un ensemble de réflexions autour de la confidentialité des données, de la dignité humaine, de l'autonomie des machines, des formes de responsabilité, de la transparence, de la justice et de l'équité, ou encore de l'impact des liens affectifs et cognitifs des relations hommes-machines sur la société et les individus. De plus, les technologies vocales intégrant de l'IA ont constitué des dispositifs concrets et adaptés pour tirer des enseignements afin d'alimenter la réflexion sur les interactions éthiques et durables.

Nous avons souhaité privilégier une vision située, systémique et développementale de l'éthique en articulant la perspective des législateurs à celle des usagers de système d'IA liée aux technologies vocales. Nous défendons ainsi l'hypothèse que la conception d'une situation et d'interactions intégrant des critères éthiques et durables contribuera à développer une acceptabilité et une appropriation du système d'IA par les utilisateurs.

L'ajout de la dimension durable à celle d'éthique permet de resituer cette dernière dans une dimension temporelle et évolutive pour considérer la relation entre l'homme et le système à base d'IA dans un processus de couplage fait d'adaptations et de transformations respectives. D'une part, ces visées éthiques et durables impliquent de prendre en compte la dimension développementale de l'humain (développement de compétences, évolution des usages et appropriation, création de ressources, développement de connaissances et d'une réflexivité, etc.). L'évolutivité du système lui permet de s'adapter aux besoins et aux contextes changeants, aux systèmes de valeurs des utilisateurs, ainsi qu'à la dimension temporelle et contextuelles des situations tout en s'améliorant intrinsèquement selon « ses » finalités propres (ou plutôt celles des concepteurs). D'autre part, cette approche considère la relation entre l'utilisateur et le système comme asymétrique : elle donne un primat au point de vue de l'acteur humain et considère que la relation avec le système à base d'intelligence artificielle doit permettre un développement des acteurs (humain et système) uniquement dans des conditions acceptables par l'humain et dans une logique de complémentarité. Sont ainsi posées les questions d'utilité de l'IA et du développement de moyens d'agir au service de l'homme, questions faisant écho aux débats éthiques sous-tendant le remplacement de l'homme par des systèmes dits « intelligents ».

Aborder l'éthique à partir d'un angle situé et systémique place inévitablement le projecteur sur l'activité de l'humain impliquant les compromis et les activités qu'il réalise en situation, qu'il s'empêche de faire ou qu'il souhaiterait mener selon ses propres motifs et valeurs intrinsèques, ou encore, les ressources qu'il déploie pour

surmonter les contraintes d'usage ou le « manque d'intelligence » (intelligence de contexte, de l'implicite, de coopération, de modes dégradés et développementale), de ces assistants dits « intelligents ». La mise à jour des activités réelles des utilisateurs, de leurs préoccupations, et de leurs logiques intrinsèques couplée à la vision de l'éthique portée par les législateurs nous a permis de construire une vision plus globale et réaliste de l'éthique pour l'inscrire dans une logique durable d'émergence de nouvelles ressources (humaines et systèmes) et de complémentarité.

Nous considérons l'éthique comme un outil au service de l'homme, de son bien-être, du bon fonctionnement de la société et du vivre-ensemble. Loin de se figer définitivement, l'éthique tend elle aussi à se développer et à s'actualiser au fil du temps, des progrès technologiques et des usages en situation. En progression constante⁴, l'éthique évolue au même titre que les avancées technologiques questionnant sans cesse les moyens à mettre en œuvre pour protéger les valeurs humaines, tout en privilégiant le bien agir et la prudence face à l'opacité et la complexité de certains enjeux associés à l'IA. Cet outil au service de l'homme impacte inévitablement son environnement, les dispositifs qui l'entoure et sa propre activité dans les sphères professionnelles et privées. Pour ces raisons une éthique située et durable étudiée et intégrée en amont de la conception permettrait d'encourager des formes d'interactions plus efficaces entre les hommes et les systèmes d'IA.

Ces questions ouvrent un ensemble de perspectives à poursuivre, en particulier vers l'étude des formes d'interaction entre l'humain et les systèmes d'IA en situation réelle. En complément de l'analyse de la littérature ici présentée, nous réalisons actuellement des observations en situation réelle afin de modéliser le développement et les utilisations d'écosystèmes d'objets connectés, dont des systèmes à base de reconnaissance vocale. Ces travaux de recherche en ergonomie contribuent à la conception de services et de dispositifs dans l'habitat dans le secteur énergétique adaptés au grand public et s'insérant de façon appropriée et durable dans les activités et les routines des habitants.

Bibliographie

Alliance des Sciences et Technologies du Numérique – Allistene (2014). *Éthique de la recherche en robotique*. Rapport n°1 de la CERNA, Commission de réflexion sur l'Éthique de la Recherche en sciences et technologies du Numérique d'Allistene.

Bekey G. A. (2012). Current trends in robotics: technology and ethics. In: Lin, P., Abney, K. and Bekey, G. A. eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. London, MIT Press, pp. 17-34

4. Nous considérons l'éthique comme une notion vivante et en évolution constante avec l'humain et les systèmes qui l'entourent. Pour ces raisons, nous évitons l'usage de notions telles que « normes » ou « principes éthiques » qui nous semble figé et réducteur. Nous privilégions la variété des points de vue sur l'éthique portés par un ensemble d'acteurs (concepteurs, législateurs, utilisateurs, philosophes, juristes, etc.) sur une variété de systèmes techniques et de contextes différenciés.

- Berger E., Crescentini A., Galeandro C., Mainardi Crohas G. (2010). La triangulation au service de la recherche en éducation. Exemples de recherches dans l'école obligatoire. *Actes du congrès de l'Actualité de la recherche en éducation et en formation (AREF)*.
- Bussone A., Stumpf S., O'Sullivan D. (2015). The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems. *In Proceedings of the 2015 International Conference on Healthcare Informatics (ICHI '15)*. IEEE Computer Society, Washington, DC, USA, 160-169.
- Cardon D. (2015). *A quoi rêvent les algorithmes. Nos vies à l'heure*. Le Seuil.
- Catellani A., Cobut E., Donjean C. (2017). *Vers davantage d'éthique en communication*. Edipro.
- Charney A. (2018). Amazon Alexa s'esclaffe à cause d'un problème d'écoute. *01net.com*, <http://www.01net.com/actualites/quand-amazon-alexa-s-esclaffe-et-terrie-ses-utilisateurs-1391170.html> (consulté le 02.09.2018)
- Cobb S. (2017). Amazon Echo and the Alexa dollhouses: Security tips and takeaways. *We live security.com*, <https://www.welivesecurity.com/2017/01/10/amazon-echo-alexa-dollhouse-security/> (consulté le 02.09.2018)
- Commission mondiale d'éthique des connaissances scientifiques et des technologies – COMEST (2017). *Rapport de la COMEST sur l'éthique de la robotique*. Septembre 2017.
- Commission Nationale Informatique & Libertés - CNIL (2017-a). *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*. Synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion confiée par la loi pour une république numérique, Décembre 2017.
- Commission Nationale Informatique & Libertés - CNIL (2017-b). *Pack de conformité. Véhicules connectés et données personnelles*.
- Davis B. (2016). The problem with voice user interfaces like Amazon Alexa. *Econsultancy*, <https://econsultancy.com/blog/68499-the-problem-with-voice-user-interfaces-like-amazon-alexa> (consulté le 02.09.2018)
- DeRenzi B., Dell N., Wacksman J., Lee S., Lesh N. (2017). Supporting Community Health Workers in India through Voice- and Web-Based Feedback. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA.
- Dubois M., & Bobillier-Chaumon M.-É. (2009). L'acceptabilité des technologies : bilans et nouvelles perspectives. *Le travail humain*, 72(4), 305.
- Edip A. (2017). Siri, Alexa, Google Home, M... Les assistants numériques, révolution ou gadget dangereux ? *Capital.fr, Polemik*, <https://www.capital.fr/polemik/siri-alexa-google-home-m-les-assistants-numeriques-revolution-ou-gadget-dangereux-1230739> (consulté le 02.09.2018)
- Epley N, Waytz A., Caciopo J.T. (2007). On seeing humain: A three factor theory of anthropomorphism. *Psychology Review*, vol 114, n°4, pp. 864-887, 2007.
- European Data Protection Supervisor – EDPS (2015). *Vers une nouvelle éthique numérique. Données, dignité et technologie*. Avis n°4/2015.

- Fleury C. (2017). La place de l'éthique dans nos vies individuelle et collective. Présentation à l'USI Events, Juin 2017. https://www.youtube.com/watch?v=x3u8xl_fxwk (consulté le 02.09.2018)
- Fortin M-F. (1996). *Le processus de la recherche : de la conception à la réalisation*. Montréal, Décarie éditeur.
- Frankish K., Ramsey W.M. (2014). *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press.
- Fréjus M. (À paraître). Élargissement et renouvellement des questions traitées par l'ergonomie dans le domaine du développement durable : retour sur 12 ans de travaux sur les activités domestiques et la maîtrise des consommations énergétiques. *Psychologie française*.
- Garcia-Acosta G., Saravia Pinilla M. H., I Romeva C. R. (2012). Ergoecology: evolution and challenges. *Work* 41, 2133-2140.
- Gardner-Bonneau D. (1999). *Human Factors and Voice Interactive Systems*. Kluwer Academic Publishers, Norwell, MA, 1999
- Guillaud H. (2016). Vers une éthique pour l'intelligence artificielle ? *Internet actu.net*, <http://www.internetactu.net/a-lire-ailleurs/vers-une-ethique-pour-lintelligence-artificielle/> (consulté le 02.09.2018)
- Guillaud H. (2017). L'IA à la recherche de règles éthiques. *Internet actu.net*, <http://www.internetactu.net/a-lire-ailleurs/lia-a-la-recherche-de-regles-ethiques/> (consulté le 02.09.2018)
- Harris R. (2005). *Voice interaction Design, Crafting the new conversational speech systems*. San Francisco: Morgan Kaufmann Publishers.
- Haué J.B. (2004). Intégrer les aspects situés de l'activité dans une ingénierie cognitive centrée sur la situation d'utilisation. *@ctivités*, vol. 1, n° 2, p. 170-194.
- Heritage S. (2017). Amazon's Alexa is now part of the family – I just hope she doesn't replace me. *The guardian.com*, <https://www.theguardian.com/technology/2017/nov/19/amazon-alexa-voice-recognition-2017-best-gadgets> (consulté le 02.09.2018)
- Heudin J.C. (2017). *Intelligence artificielle. Manuel de survie*. Science eBook.
- Hillard H. (2018). Amazon. L'assistant personnel Alexa se met à rire tout seul et effraye les utilisateurs. *Ouest-France.fr*, <https://www.ouest-france.fr/economie/entreprises/amazon/amazon-son-assistant-personnel-alexa-se-met-rire-tout-seul-et-effraye-les-utilisateurs-5610264> (consulté le 02.09.2018)
- IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html (consulté le 02.09.2018)
- Institut National de Recherche en Informatique et en Automatique - INRIA (2016). *Intelligence Artificielle. Les défis actuels et l'action d'Inria*. Livre blanc N°01. Coordonné par B. Braunschweig.
- ISO 9241-11 (1998). Exigences ergonomiques pour travail de bureau avec terminaux à écrans de visualisation (TEV) - Partie 11: Lignes directrices relatives à l'utilisabilité.

- Journal officiel de l'Union européenne (2016). *Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données*. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679> (consulté le 02.09.2018)
- Kermisch C. (2014). L'évaluation éthique des technologies en présence de risques et d'incertitudes : une analyse critique des matrices éthiques, *Proceedings du 19^e Congrès de Maîtrise des risques et de sûreté de fonctionnement*.
- Kermisch C. (2016). Les matrices éthiques au service des technologies à risques. Le cas de la gestion des déchets radioactifs à longue durée de vie, *Proceedings du 20^e Congrès de Maîtrise des risques et de sûreté de fonctionnement*.
- Klemmer S. R., Hartmann B., Takayama L. (2006). How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems*. ACM, 140–149.
- Kulesza T., Stumpf S., Burnett M., Yang S., Kwan I., Wong W-K (2013). Too much, too little, or just right? Ways explanations impact end users' mental models. *Proceedings of IEEE Symposium on Visual Languages and Human-Centric Computing*, pp. 3-10.
- Lagarigue J., Lebe G. (1997). Éthique ou morale ? *Recherche & Formation*, N°24, 1997. Conscience éthique et pratiques professionnelles. pp. 121-130.
- Lefevre F. (2017). Interaction vocale & intelligences artificielles : état des lieux et opportunités. *Medium.com*, <https://medium.com/@francklefevre/interaction-vocale-intelligences-artificielles-%C3%A9tat-des-lieux-et-opportunit%C3%A9s-3b45239ee4a7> (consulté le 02.09.2018)
- Luria M., Hoffman G., Zuckerman O. (2017). Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 580-628.
- Mephram B. (2000). A framework for the ethical analysis of novel foods: the ethical matrix. *Journal of agricultural and environmental ethics*, 12(2), 2000, p. 165-176.
- Metwally O., Gregory B. (2017). Alexa, start living healthy! *Workshop on Interactive Systems in Health Care (WISH)*, http://wish.gatech.edu/wp-content/uploads/metwally_omar.pdf (consulté le 02.09.2018)
- Minker W., Néel F. (2002). Développement des technologies vocales. *Le travail humain*, vol. 65, no. 3, pp. 261-287.
- Morel P.M. (2017). Vertu éthique et rationalité pratique chez Aristote. Note sur la notion d'hexis proairetiké », *Philonsorbonne*, 11/2017, 141-153.
- Munteanu C., Penn G. (2016). Speech-based Interaction: Myths, Challenges, and Opportunities. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 992-995
- Nielsen J. (1993). *Usability engineering*. Boston: Academic Press.
- Paolino J. (2017). Google Home vs Alexa. Two Simple User Experience Design Gestures that Delighted a Female User. *Medium.com*, <https://medium.com/startup-grind/google-home-vs-alexa-56e26f69ac77> (consulté le 02.09.2018)
- Pariser E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press.

- Pelletier C., Moreau É. M.-F. (2008). L'appropriation des technologies de l'Internet et ses facteurs critiques de succès : un défi de plus pour les PME ? *Revue internationale P.M.E : économie et gestion de la petite et moyenne entreprise*, vol. 21, 2, pp. 75-117.
- Pestanes P., Gautier B. (2017). Essor des assistants vocaux intelligents : nouveau gadget pour votre salon ou fenêtre d'opportunité pour rebattre les cartes de l'économie du web ?, *Wavestone*, 2017.
- Pinard R., Potvin P., Rousseau R. (2004). Le choix d'une approche méthodologique mixte de recherche en éducation. *Recherches qualitatives*, vol.24, pp.58-82.
- Portet F., Vacher M., Rossato S. (2012). Les technologies de la parole et du TALN pour l'assistance à domicile des personnes âgées : un rapide tour d'horizon, *Atelier ILADI: Interactions Langagières pour personnes Agées Dans les habitats Intelligents*, Grenoble.
- Portolan N., Nael M., Renoullin JL, Naudin S. (1999). Will we speak to our TV remote control in the future? In: *Proceedings of the 17th international symposium on human factors in telecommunication, HFT'99*, Copenhagen.
- Proulx S. (2002). Trajectoires d'usages des technologies de communication : les formes d'appropriation d'une culture numérique comme enjeu d'une société du savoir. *Annales des télécommunications*, 57(3-4), 180-189.
- Rapport parlementaire (2017). *Pour une intelligence artificielle maîtrisée, utile et démystifiée*. Rapporteurs : C. de Ganay & D. Gillot, fait au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques. N° 464 tome I (2016-2017), Mars 2017.
- Rasmus D. (2016). Hey Alexa: The Pros and Cons of My Amazon Echo Experience. *Iphonelife.com*, <https://www.iphonelife.com/content/hey-alexa-pros-and-cons-my-amazon-echo-experience> (consulté le 02.09.2018)
- Rodin J. (1986). Aging and health: effects of the sense of control. *Science*, 233(4770):1271-1276.
- Rouvroy A., Berns T. (2013). Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ?, *Réseaux* 2013/1 (n° 177), p. 163-196.
- Russel S.J., Norvig P. (2009). *Artificial intelligence: a modern approach*. Pearson.
- Shackel, B. (1991). Human Factors for Informatics Usability. In B. Shackel & B. Richardson (Eds.), *Human Factors* (pp. 21-38). Cambridge: Cambridge University Press.
- Soronen, H., Turunen, M., and Hakulinen, J. (2008). Voice Commands in Home Environment - a Consumer Survey. In *Proceedings of Interspeech 2008*, 2078-2081.
- Stumpf, S. (2015). The role of explanations in interacting with intelligent machines. Seminar. https://www-users.cs.york.ac.uk/alistair/HCISeminars/York_seminar_Nov15.pdf (consulté le 02.09.2018)
- Tisseron S. (2015). *Le jour où mon robot m'aimera. Vers l'empathie artificielle*. Albin Michel.
- Tual M. (2017). L'intelligence artificielle reproduit aussi le sexisme et le racisme des humains. *Le Monde.fr*, http://www.lemonde.fr/pixels/article/2017/04/15/quand-l-intelligence-artificielle-reproduit-le-sexisme-et-le-racisme-deshumains_5111646_4408996.html (consulté le 02.09.2018)
- Turckle S. (2015). *Seuls ensemble*. Paris, L'Échappée.

- Villani C., Schoenauer M., Bonnet Y., Berthet C., Cornut A-C., Levin F. & Rondepierre B. (2018). Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne. Mission parlementaire. https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf (consulté le 03.09.2018).
- Whitenton K. (2016). Voice Interaction UX: Brave New World...Same Old Story. *Nielsen Norman Group.com*, <https://www.nngroup.com/articles/voice-interaction-ux/> (consulté le 02.09.2018).

