Agents conversationnels: Enjeux d'éthique

Comité National Pilote d'Éthique Du Numérique Collectif, Alexei Grinbaum, Laurence Devillers, Gilles Adda, Raja Chatila, Caroline Martin, Célia Zolynski, Serena Villata

AVIS N°3
AGENTS CONVERSATIONNELS : ENJEUX D'ÉTHIQUE

COMITÉ NATIONAL PILOTE
D'ÉTHIQUE DU NUMÉRIQUE

sous l'égide du
COMITÉ CONSULTATIF NATIONAL D'ÉTHIQUE
POUR LES SCIENCES DE LA VIE ET DE LA SANTÉ

AVIS N°3

AGENTS CONVERSATIONNELS : ENJEUX D'ÉTHIQUE

AVIS ADOPTÉ LE 15 SEPTEMBRE 2021 À L'UNANIMITÉ DES MEMBRES PRÉSENTS LORS DE L'ASSEMBLÉE PLÉNIÈRE DU CNPEN

SAISINE DU PREMIER MINISTRE

Dans sa lettre du 15 juillet 2019 donnant mission au président du CCNE de mettre en œuvre une démarche pilote concernant les questions d'éthique des sciences, technologies, usages et innovations du numérique et de l'intelligence artificielle, le Premier ministre a souhaité que les travaux conduits dans cette phase pilote concernent en particulier le diagnostic médical et l'intelligence artificielle, les agents conversationnels ainsi que le véhicule autonome. Cet avis du Comité national pilote d'éthique du numérique (CNPEN) porte sur les agents conversationnels.

SOMMAIRE

I. INTRODUCTION

Un agent conversationnel (appelé aussi chatbot[1]) est une machine qui, à travers des échanges écrits ou oraux, interagit avec son utilisateur en langage naturel. Le plus souvent, un agent conversationnel ne constitue pas une entité indépendante mais est intégré dans un système ou une plateforme numérique multitâche, comme un smartphone ou un robot.

Parmi les travaux en éthique de l'intelligence artificielle, la réflexion sur les agents conversationnels se distingue en premier lieu par la place du langage dans ces systèmes ; cela veut aussi bien dire l'analyse de l'impact des systèmes d'apprentissage machine sur le langage humain que l'impact du langage tel qu'il est utilisé par ces systèmes, sur les utilisateurs et la société en général. Alors qu'il n'existe actuellement que peu d'études consacrées à ces questions, cet avis vise à éclairer les enjeux et les défis qu'induit le déploiement des agents conversationnels à grande échelle.

Comme pour tous les systèmes d'intelligence artificielle, cette réflexion, en s'appuyant sur les valeurs éthiques, met en lumière des principes de conception et des préconisations qui seront énumérés dans cet avis. L'UNESCO souligne[2] que la prise en compte des risques et des préoccupations éthiques ne devrait pas entraver l'innovation et le développement, mais plutôt offrir de nouvelles possibilités technologiques et stimuler la recherche et l'innovation, ainsi que la réflexion morale. Même si chacune des valeurs éthiques et chacun des principes de conception sont désirables, toute situation concrète peut faire émerger des tensions entre eux, par exemple entre la sécurité publique et la liberté individuelle, ou encore entre l'efficacité et la transparence des systèmes d'intelligence artificielle. La prise de décision au cas par cas est alors nécessaire en tenant compte des contextes de conception et d'usage, dans le respect du principe de proportionnalité et des droits fondamentaux. Dans chaque cas concret, la délibération doit s'appuyer autant sur les finalités recherchées que sur les contraintes techniques, ainsi que sur la prise en compte des intérêts des utilisateurs à court et à long terme.

Des chatbots capables de dialogue vocal ou écrit sont déjà déployés dans le domaine de la santé, de l'aide aux personnes vulnérables, du recrutement, du service après-vente, de l'éducation, des banques, des assurances et dans bien d'autres encore, pour rendre de nombreux services aux utilisateurs. Dans le domaine de la santé, par exemple, les agents conversationnels sont utilisés pour le diagnostic, la surveillance ou encore l'assistance aux patients. Le déploiement des chatbots dans les entreprises vise à supprimer des tâches répétitives, améliorer l'interaction avec les clients et réduire les coûts. Le déploiement des agents conversationnels peut aussi avoir des visées éducatives ou ludiques.

Dans la majorité des cas, les chatbots actuels répondent selon des stratégies prédéterminées par leurs concepteurs. Du point de vue de l'utilisateur, cette solution prédéterminée est limitée car elle donne l'impression que « l'agent conversationnel manque d'imagination ». Le succès d'une telle stratégie dans des dialogues complexes, ainsi que la capacité d'un agent conversationnel de ce type à expliquer son comportement, sont ainsi restreints. Ce sont toutefois des facteurs déterminants pour la diffusion de cette technologie. La situation est en train de changer avec le développement des chatbots qui utilisent les modèles de langue capables de construire des dialogues plus réalistes. À l'heure actuelle, les concepteurs des agents conversationnels cherchent à créer des systèmes personnalisés pour qu'ils engagent l'utilisateur de manière plus efficace. Les recherches scientifiques et technologiques sur les agents conversationnels sont motivées par des visions ambitieuses : un « ami virtuel » imitant des affects et capable d'apprendre en interaction avec l'utilisateur, ou encore un « ange gardien » qui veillera à la sécurité de nos données personnelles. Ces recherches s'appuient sur des technologies de pointe dans le domaine de l'apprentissage machine, développées par les instituts de recherche à l'international et diffusées principalement par les géants du numérique, telles que les réseaux de neurones de type transformeur, nourris par de gigantesques collections de données. Ces outils ont récemment amplifié le champ des possibles en matière de reconnaissance de la parole et de génération automatique de textes. Les chatbots les plus récents soulèvent de multiples questions éthiques également liées à l'utilisation de l'informatique affective[3] qui permet d'influencer le comportement des utilisateurs.

COMMENT CONSTRUIT-ON UN CHATBOT ?

Les agents conversationnels ont longtemps été construits avec une architecture modulaire utilisant des technologies du traitement automatique de la langue (TAL). Ces modules reposent sur des algorithmes d'apprentissage machine ou suivent des règles décidées et transposées dans le code par des concepteurs humains. Un agent conversationnel de ce type, par exemple, un assistant vocal, comprend des modules d'analyse du signal, de reconnaissance de la parole, de traitement sémantique, de gestion des stratégies et de l'historique du dialogue, d'accès à des connaissances internes ou externes (bases de connaissances publiques ou spécifiques, ontologies, données disponibles sur le web), de génération de réponses et de synthèse de la parole. Ces dernières années, développer soi-même un chatbot rudimentaire, écrit ou oral, est devenu relativement facile grâce à la disponibilité de nombreux outils de conception[4].

Les nouvelles générations de chatbots sont de plus en plus performantes grâce à l'évolution des techniques d'apprentissage machine, à la puissance des processeurs et à la taille des bases de données. Les modèles de langue, c'est-à-dire des modèles qui prédisent le mot ou la séquence susceptible d'être pertinent dans un contexte de discours, sont devenus « le graal » du développement des applications en TAL, et en particulier des chatbots. Les modèles les plus récents sont appelés des « transformeurs » : ce sont des réseaux de neurones qui, à partir de vastes corpus linguistiques, apprennent les régularités les plus saillantes, sans être influencé par l'ordre des mots.

[1] Ces termes ont parfois reçu des significations différentes dans le passé (un chatbot serait un système qui n'interagit que par écrit et ne possède pas de mémoire). Dans le présent avis, « agent conversationnel » et « chatbot » sont traités comme synonymes, ce qui correspond à l'évolution actuelle des technologies.

[2] UNESCO, Projet de recommandation sur l'éthique de l'intelligence artificielle, 25 juin 2021. https://unesdoc.unesco.org/ark:/48223/pf0000377897_fre

[3] L'informatique affective est le développement de systèmes dotés de capacités de reconnaître, d'exprimer, de synthétiser et modéliser les émotions humaines.

[4] Par exemple, LiveEngage, Chatbot builder, Passage.ai, Plato Research Dialogue System.

COMMENT CONSTRUIT-ON UN CHATBOT ?

Cette technique a été développée à partir de 2017. Depuis la mise à disposition du modèle de langage BERT (Bidirectional Encoder Representations from Transformers) par Google[5], de nombreuses avancées scientifiques ont vu le jour. On peut citer en juillet 2020 GPT-3 d'OpenAI avec 175 milliards de paramètres et 570 gigaoctets de données d'apprentissage ou LaMDA (Language Model for Dialogue Applications) de Google, entraîné spécifiquement sur des données conversationnelles qui permettent d'engager un dialogue libre sur un nombre potentiellement infini de thèmes. Plus récemment, l'on a recensé : début 2021, le modèle Switch-C, également développé par Google, avec 1600 milliards de paramètres : en été 2021, le modèle Jurassic-1 Jumbo de AI21 Labs (Israël) avec 178 milliards de paramètres, YaML de Yandex (Russie) avec 13 milliards de paramètres pour la langue russe ou encore WuDao 2.0 de BAAI (Chine) avec 1750 milliards de paramètres, dont le modèle est orienté vers l'anglais et le mandarin. A notre connaissance, WuDao est désormais le plus grand réseau de neurones artificiels jamais créé. Ces modèles ne sont souvent pas disponibles en accès ouvert et contiennent de multiples biais implicites, notamment liés au caractère opaque des données utilisées pour l'apprentissage. Dans le souci de transparence, l'initiative Big Science, portée par Huggingface et un consortium de laboratoires de recherche publique, vise à créer un modèle multilingue de taille équivalente, ouvert et accessible, qui permettra de mieux comprendre le fonctionnement et les limitations de ces gigantesques transformeurs.

La réalité technologique et les connaissances dans le domaine des agents conversationnels évoluent rapidement. Le passage aux algorithmes d'apprentissage profond, à l'apprentissage adaptatif ou encore aux transformeurs créent de nouvelles tensions éthiques pour lesquelles les ingénieurs et les concepteurs doivent proposer des solutions conformes aux normes et valeurs de la société. Mais les normes et valeurs évoluent aussi, en particulier sous l'influence des technologies numériques. Ces technologies augmentent non seulement la difficulté de distinguer le dialogue artificiel de la parole humaine, mais aussi l'impact cognitif et affectif sur les utilisateurs humains. Lorsque les agents conversationnels utilisent le langage humain, l'anthropomorphisation des chatbots par leurs utilisateurs est une tendance naturelle : le perfectionnement de ces technologies tend à brouiller la frontière perçue entre les machines et les êtres humains.

Dans le but de nourrir sa réflexion, le CNPEN a recueilli les avis des citoyens et des parties prenantes via un appel à contribution sur les tensions éthiques relatives aux agents conversationnels, présenté dans l'annexe 2. Ces avis sont partagés et nettement polarisés. Certains montrent une forte anthropomorphisation de l'agent conversationnel qui pousse les contributeurs à le mettre sur un pied d'égalité avec une personne humaine ; d'autres, au contraire, ont tendance à le réduire à un outil automatique anodin. Les réponses se rejoignent cependant sur le fait que les enjeux et les jugements éthiques dépendent des finalités, des applications concrètes et des personnes concernées par les contextes d'usage. Les questions juridiques occupent également une place importante dans les réponses des contributeurs. Celles qui sont spécifiques aux agents conversationnels, sont présentées tout au long de cet avis ainsi que dans l'annexe 1, relative aux enjeux du consentement et de la protection des données personnelles.

Les tensions éthiques que soulèvent les agents conversationnels appellent un développement réfléchi et responsable de ces systèmes. La question de la responsabilité est posée dans toutes ses formes : responsabilité légale et morale, individuelle et collective, celle du concepteur, du fabricant, de l'utilisateur et du décideur politique, celle relative aux éventuels dysfonctionnements et celle liée aux conséquences de ces technologies à long terme. Certains enjeux font d'ores et déjà l'objet d'une réglementation. Ainsi, la CNIL a identifié dans son livre blanc sur les assistants vocaux, paru en septembre 2020, des questions qui s'insèrent dans le cadre du Règlement général sur la protection des données (RGPD)[6].

Au-delà de ces points réglementaires, se pose la question du sens de la relation humain-machine et des responsabilités qu'elle induit. Quels comportements et croyances avons-nous par rapport aux agents conversationnels ? Quel comportement le concepteur doit-il donner au chatbot, celui-ci doit-il imiter systématiquement l'être humain ? Un chatbot peut-il mentir à son utilisateur ? Les erreurs éventuelles des agents conversationnels seront-elles plus ou moins acceptables que celles d'un être humain ? Quelles sont les limites de cette comparaison ?

Le langage humain conditionne, voire détermine, les spécificités culturelles[7], les perceptions ou encore les visions du monde. Dans le cas des agents conversationnels, ces représentations langagières sont dépourvues d'expérience vécue. Les répliques, formulées par un système de génération de la parole qui ne peut percevoir physiquement, ni avoir de sentiments, ni raisonner de manière humaine, conduisent à créer un univers linguistique sans corporalité et sans compréhension du sens. L'emploi du langage par les agents conversationnels supprime ainsi le lien univoque entre le langage et l'être humain : nous échangeons des répliques avec des machines qui ne peuvent ni assumer ce qu'elles disent, ni en être responsables.

Cependant, le langage de l'agent conversationnel est susceptible d'influencer la pensée de son utilisateur en y imprimant des notions, des perceptions, des idées ou encore des croyances. L'utilisateur bâtit alors un monde où le langage des machines s'ajoute au réel et s'intègre dans son environnement social. Ce monde aux frontières recomposées est de plus en plus présent à l'individu et transforme subrepticement le sens de ses valeurs, comme par exemple l'autonomie et la dignité de l'utilisateur devant un agent conversationnel capable de lui mentir ou le manipuler. À l'échelle de la société, la question des discriminations induites par les agents conversationnels doit être traitée dans le souci d'équité. À long terme, les effets des chatbots, par exemple les « deadbots »[8], risquent de produire une évolution sensible de la condition humaine. La co-adaptation langagière entre les utilisateurs humains et les agents conversationnels est le moteur de cette transformation.

[5] https://blog.google/technology/ai/lamda

[6] https://www.cnil.fr/fr/votre-ecoute-la-cnil-publie-son-livre-blanc-sur-les-assistants-vocaux

[7] « La culture, dans son sens le plus large, est considérée comme l'ensemble des traits distinctifs, spirituels et matériels, intellectuels et affectifs, qui caractérisent une société ou un groupe social. Elle englobe, outre les arts et les lettres, les modes de vie, les droits fondamentaux de l'être humain, les systèmes de valeurs, les traditions et les croyances.» Déclaration de Mexico sur les politiques culturelles. Conférence mondiale sur les politiques culturelles, Mexico City, 26 juillet - 6 août 1982.

[8] On appelle « deadbot » un agent conversationnel imitant à dessein la manière de parler ou d'écrire d'une personne décédée.

II. QUESTIONS ÉTHIQUES RELATIVES AUX USAGES DES CHATBOTS

Les agents conversationnels sont de plus en plus intégrés dans différents aspects de la vie humaine. Leur utilisation soulève des tensions éthiques, ce qui pose la question de la responsabilité au sens de la philosophie morale. Celle-ci est – et doit être – du ressort des êtres humains car la machine n'est pas un agent moral et ne doit en aucun cas être considérée comme une personne. La responsabilité est ainsi partagée entre l'utilisateur d'un agent conversationnel, le concepteur (informaticien ou groupe d'informaticiens possédant des connaissances spécialisées), l'entraîneur (individu ou groupe d'individus effectuant la sélection et le tri des données et l'optimisation du système d'apprentissage machine) et le fabricant (personne physique ou morale qui met un chatbot sur le marché). Ce partage s'effectue au cas par cas, en fonction des aspects techniques et de l'implication de l'utilisateur, du concepteur et du fabricant dans chacune des situations qui provoquent des tensions éthiques.

La conception et l'utilisation des agents conversationnels actuellement déployés et de ceux qui le seront demain doit être interrogée à l'aune des questions éthiques. Ainsi cet avis s'adresse-t-il aux chercheurs en informatique qui doivent s'interroger sur leurs méthodologies de conception et d'évaluation des agents conversationnels. Il s'adresse également aux industriels qui doivent prendre conscience des tensions d'éthique et de confiance, avec les conséquences qu'elles induisent sur le marché économique, et doivent soutenir des travaux permettant de les lever. Enfin, cet avis s'adresse aussi aux autorités publiques qui doivent avoir la responsabilité d'accroître l'effort de formation et d'éducation, mais aussi d'évaluation des effets des agents conversationnels à court terme et des expérimentations à l'échelle de la société permettant de saisir leurs effets à long terme.

L'approche de toutes ces parties prenantes doit se fonder sur la transparence et l'explicabilité pour garantir le respect des valeurs telles que l'autonomie humaine, la dignité et l'équité. L'ensemble de leurs actions doit également correspondre au cadre « éthique par conception » préconisé par l'Union européenne.

Le CNPEN a dégagé huit dimensions de réflexion éthique relatives aux usages des chatbots (chapitre II) et cinq dimensions relatives aux technologies (chapitre III), qui motivent les treize préconisations sur les usages, les dix principes de conception des chatbots et les onze questions de recherche formulés dans cet avis. Parmi ces différents enjeux, les trois principales tensions concernent le statut des agents conversationnels, l'imitation du langage et des émotions par les chatbots et la prise de conscience par le public des capacités et limitations des agents conversationnels, y compris leur capacité à manipuler.

La réalité technologique et les connaissances dans le domaine des agents conversationnels évoluent très rapidement. De ce fait, la réflexion sur les enjeux d'éthique des agents conversationnels évoluera nécessairement pour prendre en compte les évolutions culturelles et technologiques induites. Cette réflexion devra être poursuivie à un horizon de trois à cinq ans.

N'étant pas spécialistes en informatique, les utilisateurs ne connaissent que peu, voire pas du tout, les capacités des chatbots. De plus, certaines capacités des agents conversationnels, encore au stade de développement et peu mises en œuvre, sont survendues par le marketing des technologies. Les croyances des utilisateurs sont souvent nourries par la fiction. Leur responsabilité peut toutefois être engagée dans certaines situations, notamment en cas d'usage malveillant ou maladroit.

Quant aux concepteurs des agents conversationnels, ils n'ont souvent pas conscience des tensions éthiques qui peuvent émerger pendant l'utilisation des chatbots. Cela est lié soit à l'impossibilité de prévoir les conséquences de cette technologie, soit au manque de vigilance ou d'expérimentation lors de la conception. Cependant, la responsabilité des concepteurs comme celle des fabricants est engagée dans tous les cas.

HISTOIRE

Le premier agent conversationnel de l'histoire de l'informatique est le programme ELIZA de Joseph Weizenbaum (1966) au MIT, qui est aussi l'un des premiers leurres conversationnels. ELIZA, qui joue le rôle d'un psychothérapeute rogérien[11], simule un dialogue écrit en réutilisant les mots ou répliques de l'utilisateur « patient » dans ses réponses. Si une phrase de l'utilisateur contient le mot « ordinateur », Eliza demande par exemple : « Tu dis ça parce que je suis une machine ? », ce qui peut prêter à confusion. Aujourd'hui, l'expression « effet ELIZA » désigne la tendance à assimiler de manière inconsciente le dialogue avec un ordinateur à celui avec un être humain.

[9] E Ruane, A Birhane, A Ventresque, Conversational AI: Social and Ethical Considerations.AICS, 104-115, 2019.

[10] https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

[11] La Psychothérapie Rogérienne est fondée sur l'idée que le patient a les solutions et les ressources nécessaires en lui pour résoudre ses problèmes. Le thérapeute, par le dialogue, l'aide à développer ses propres choix sans l'orienter ou l'influencer.

HISTOIRE

L'histoire des agents conversationnels prend ses origines dans le jeu de l'imitation d'Alan Turing (Turing, 1950). Turing propose un test d'intelligence pour une machine à travers la capacité à dialoguer en langage naturel. Dès 1991, des concours sont organisés afin de soutenir le développement de chatbots capables de réussir le test de Turing. Ce test est un jeu d'imitation dans lequel une entité cachée, qui peut être un expérimentateur humain ou une machine, interagit par écrit avec un sujet humain. L'objectif pour l'entité cachée est de faire en sorte que son interlocuteur pense qu'elle est également humaine. Alan Turing prédit ainsi qu'en 2000, les machines seront capables de tromper environ 30 % des juges humains durant un test de cinq minutes. Il existe une compétition annuelle créée en 1990, le prix Loebner, récompensant le programme considéré comme le plus proche de réussir le test de Turing. Malgré les avancées certaines de la technologie, ce test apparaît insuffisant pour capter toute la complexité de l'intelligence humaine. D'autres tests sont proposés pour le compléter, par exemple « le test de Lovelace »[12], dans lequel un agent artificiel réussit une tâche créative seulement si son programmeur ne peut pas expliquer comment le texte produit par la machine a été obtenu. Ce critère de l'intelligence a été également critiqué. Tous ces tests sont formulés avec un critère négatif : si une machine ne les réussit pas, alors elle n'est pas intelligente ; mais si elle les réussit, cela ne la rend pas encore intelligente[13]. La tendance est au découplage entre la notion d'intelligence et la résolution de problèmes.

De manière de plus en plus fréquente, les fabricants cherchent à renforcer l'impression pour l'utilisateur humain d'être en train de discuter avec un « personnage virtuel » doté d'une intelligence. Que la personnalisation soit programmée par le concepteur ou pas, l'utilisateur projette des caractéristiques humaines sur le chatbot de manière souvent spontanée et inconsciente. L'inévitabilité de cette projection ainsi que sa portée anthropologique, psychologique, juridique et politique suscitent de multiples interrogations éthiques. Les réponses, souvent contradictoires, à l'appel à contribution relatives aux usages des chatbots démontrent la complexité et la richesse de ces interrogations (Annexe 2).

1. STATUT DES AGENTS CONVERSATIONNELS

La parole non-humaine, depuis les oracles de l'Antiquité, est perçue comme source de révélations et de fascinations. Quelle que soit la nature de son interlocuteur – que celui-ci soit réel ou imaginaire, qu'il soit une statue, une pierre, un dieu, un animal ou une machine – l'homme projette sur lui spontanément des traits qui sont humains : genre, pensée, volonté, désir, conscience, représentation du monde. Le temps d'une conversation, cet interlocuteur apparaît alors comme un individu doté de caractéristiques apparemment familières, même si, sur le plan ontologique, il s'agit d'une existence non-humaine ou virtuelle. Cette projection met en jeu plusieurs valeurs et principes : autonomie et liberté humaines, dignité, responsabilité, loyauté, non-discrimination, justice, sécurité, respect de la vie privée.

La projection des traits humains sur un agent conversationnel est spontanée. Elle est d'abord vécue comme une brève illusion, mais elle peut aussi persister[14]. De plus, elle peut être renforcée techniquement à travers la personnification des agents conversationnels, par exemple par le timbre de la voix ou la façon de parler. Les effets de cette projection dépendent des connaissances que l'interlocuteur humain possède sur la technique, de l'état d'esprit dans lequel il se trouve et de sa disposition affective, mais aussi du degré de personnification de l'agent conversationnel choisi par le concepteur. Une information claire et compréhensible portant sur le statut du chatbot permet de limiter cette projection, sans la supprimer : l'utilisateur sera plus rapidement et plus aisément confronté à l'idée qu'il interagit avec une machine. Mais, de manière ludique ou sérieuse, l'utilisateur s'engage souvent dans un dialogue émotionnel avec un agent conversationnel tout en sachant que celui-ci est une machine. Cela démontre l'insuffisance de la seule information de l'utilisateur : il s'agit d'un réel effacement des distinctions de statut, source d'importantes tensions éthiques liées par exemple à la dignité humaine ou à la manipulation.

La différence de statut entre un agent conversationnel et un être humain est particulièrement importante sur le plan de la finalité ou du rôle attribué au chatbot. Un système informatique est conçu pour réaliser un objectif décidé par son concepteur, tandis qu'un être humain est libre de se donner ses propres objectifs ou de parler sans se poser cette question. Sur le plan éthique, le maintien des distinctions de statut ou, au contraire, leur effacement, doivent être appréciés en fonction des finalités recherchées. Dans certains contextes, l'anthropomorphisme peut engendrer une confusion néfaste ; dans d'autres, malgré les apparences, il se révélera utile pour l'utilisateur.

Selon les finalités, il peut s'avérer nécessaire d'éviter la confusion entre agent conversationnel et interlocuteur humain lors du dialogue, ainsi que d'éviter les effets néfastes qu'elle est susceptible d'entraîner.

La différence de statut s'apprécie également selon le type d'applications. Par exemple, des applications des agents conversationnels dans le domaine de la santé portent sur le conseil médical, le traitement et le diagnostic en psychiatrie ou encore en psychologie. Un « médecin virtuel » serait capable de faire le diagnostic et de préconiser le traitement de maladies communes ; un « infirmier virtuel » peut surveiller des patients. Pour la mise en place d'un protocole médical, certains chatbots s'appuient sur la projection spontanée de traits humains sur l'agent conversationnel ; d'autres sont conçus dans l'idée d'induire un effet bénéfique par le caractère explicitement non-humain du dialogue.

[12] Bringsjord, S., Bello, P. & Ferrucci, D. Creativity, the Turing Test, and the (Better) Lovelace Test. Minds and Machines 11, 3–27 (2001).

[13] Floridi, L., Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. Minds & Machines 30, 681–694 (2020).

[14] Ruane E., Farrell S., Ventresque A. (2021) User Perception of Text-Based Chatbot Personality. In: Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2020. Lecture Notes in Computer Science, vol 12604. Springer, Cham. https://doi.org/10.1007/978-3-030-68288-0

Dans le domaine de la justice, les chatbots sont de plus en plus utilisés comme aide à la décision tant pour les professionnels du droit que pour le public en permettant aux utilisateurs d'accéder à des bases de données comportant de multiples jugements ou précédents juridiques. Les agents conversationnels fournissent des solutions fondées sur l'analyse de statistiques légales (justice prédictive) ou résultant de données qui leur seraient fournies sur une affaire. Dans cette dernière hypothèse, ils se comportent en quelque sorte comme des juges virtuels (justice simulative). Il est peu probable que, dans ce domaine, puissent surgir des risques d'assimilation de la machine à une personne. En revanche, lorsque des collectivités publiques ou des opérateurs privés mettront à la disposition du public des agents conversationnels pour délivrer des informations juridiques ou faciliter la résolution de litiges, ils devront être attentifs à ce que le langage employé et, le cas échéant, l'expression vocale soient suffisamment éloignés d'une conversation humaine pour que l'agent conversationnel ne soit pas confondu avec une personne.

LES JUGES VIRTUELS

Certains États ou communautés envisagent la création de juges virtuels dans les prétoires, eux-mêmes devenus virtuels. Cette perspective peut sembler lointaine, compte tenu des difficultés techniques. Elle apparaît impossible au sein de l'Union européenne compte tenu de l'article 22 du RGPD, même si un État l'a envisagé[15]. Elle serait exclue en France sur le fondement de ce même texte et en application des dispositions des articles 47 et 120 de la loi du 6 janvier 1978 dite loi informatique et libertés, une garantie humaine étant nécessaire pour que se forme une décision de justice. Il existe, toutefois, des exemples chinois[16] et canadien. En revanche, la question de la création d'assistant virtuel du juge pourrait se poser dans l'hypothèse où des États souhaiteraient procéder à des audiences préparatoires à l'aide d'un agent conversationnel afin de réunir des réponses à certaines questions. Serait-il conçu de telle sorte qu'il ne puisse être assimilé à un juge humain, ou au contraire, s'il devait ressembler à un humain, comment seraient définies son apparence et son expression orale ?

Le CNPEN souligne qu'un chatbot ne devrait en aucun cas être perçu par l'utilisateur comme une personne responsable, même par projection. Sur le plan général, il ne s'agit ni de laisser libre cours à l'anthropomorphisation, ni de vouloir l'éliminer à tout prix, mais d'en définir des limites dans des cas concrets. Une anthropomorphisation s'étendant jusqu'à la sphère de la responsabilité représente pour la société un risque majeur d'apparition d'« agents sans foi ni loi », nouveaux et hors de contrôle. Par conséquent, il faut veiller continûment, avec vigilance et lucidité, à ce que le développement et la diffusion des « personnages virtuels » permettent d'éviter leur interprétation en tant qu'agents responsables. Cela peut éventuellement déboucher sur une mesure d'ordre réglementaire.

PRÉCONISATION 1

RÉDUIRE LA PROJECTION DE QUALITÉS MORALES SUR UN AGENT CONVERSATIONNEL

Pour réduire la projection spontanée de qualités morales sur l'agent conversationnel et l'attribution d'une responsabilité à ce système, le fabricant doit limiter sa personnification et informer l'utilisateur des biais éventuels issus de l'anthropomorphisation de l'agent conversationnel.

LA LOI CALIFORNIENNE RELATIVE À LA CONCEPTION DES CHATBOTS

À l'issue d'un débat qui a fortement mobilisé les grandes plateformes, opposées à une loi qui imposerait à tous les concepteurs d'agents conversationnels de révéler leur caractère artificiel, la Californie a adopté le « *Bolstering Online Transparency Act* » (California Senate bill 1001). Elle oblige les concepteurs des agents conversationnels utilisés pour vendre un produit ou convaincre un électeur à révéler aux interlocuteurs de leurs chatbots qu'ils dialoguent avec une machine. Il n'existe actuellement aucun texte comparable en droit français. L'article 52 de la proposition de règlement européen sur l'intelligence artificielle prévoit que les fournisseurs des agents conversationnels doivent, sous peine d'amende, s'assurer que leurs utilisateurs seront informés qu'ils communiquent avec un système d'intelligence artificielle[17]. Il est prévu que cette obligation ne s'appliquera pas dans quelques cas particuliers, par exemple pour les chatbots qui aident à détecter, prévenir, enquêter et poursuivre les infractions pénales.

PRÉCONISATION 2

AFFIRMER LE STATUT DES AGENTS CONVERSATIONNELS

Toute personne qui communique avec un agent conversationnel doit être informée de manière adaptée, claire et compréhensible du fait qu'elle dialogue avec une machine. Le format et la temporalité de cette communication doivent être adaptés au cas par cas.

[15] https://e-estonia.com/artificial-intelligence-as-the-new-reality-of-e-justice/
https://www.alexsei.com/

[16] Chris Young, China has Unveiled an AI Judge that Will 'Help' With Court Proceedings, Interesting Engineering (Aug 19, 2019) available at https://interestingengineering.com/china-has-unveiled-an-ai-judge-that-will-help-with-court-proceeding.

[17] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

2. IDENTITÉ DES AGENTS CONVERSATIONNELS

Un nom propre peut être donné à l'agent conversationnel soit par son concepteur, soit par l'utilisateur.

Dans le premier cas, nommer la machine peut aider à mieux réaliser sa fonction, par exemple dans les secteurs de service après-vente, d'assistance aux personnes ou de divertissement. Dans ces secteurs, le nom choisi par le concepteur représente la marque ou l'entreprise, avec pour but de renforcer le lien entre l'utilisateur et le fabricant du produit. En revanche, une autorité publique qui utilise un chatbot pour rendre des services peut volontairement choisir ne pas lui donner de nom, par exemple pour renforcer la perception du caractère impersonnel du pouvoir qu'elle exerce.

Dans le second cas, l'utilisateur d'un agent conversationnel lui donne un nom comme on le fait souvent avec les jouets. En individualisant l'objet, cette attribution de nom permet à l'utilisateur de constituer pour lui-même un « personnage virtuel », souvent perçu comme un « ami » ou un « partenaire ». Le choix du nom relève alors d'une liberté donnée à l'utilisateur humain, qui peut nommer « son » chatbot selon sa convenance.

À travers la possibilité de choisir le nom du chatbot, de l'effacer ou d'en choisir un nouveau, l'utilisateur acquiert une illusion de contrôle et de maîtrise vis-à-vis du « personnage virtuel ». Explicitement ou implicitement, il se perçoit souvent comme co-créateur d'un « individu » unique. Dans certains contextes d'usage, par exemple pour les « compagnons virtuels », même s'il ne s'agit que d'illusion, l'utilisateur peut percevoir le chatbot comme une entité autonome, à un niveau en quelque sorte analogue à celui des animaux domestiques.

Nommer possède plusieurs facettes : la commodité du point de vue de l'utilisateur et une illusion d'autonomie du chatbot, mais aussi une longue histoire culturelle. Certains exégèses des récits mythologiques présentent Adam comme co-créateur de la Nature à travers son action de nommer tous les êtres vivants, ou encore Prométhée comme créateur du langage. Les interprétations de ces récits en tirent des enseignements éthiques et anthropologiques. Le récit de la tour de Babel, par exemple, sans condamner l'homme pour sa propre ambition de créer, mettrait en lumière la démesure de son action, ouvrant ainsi un débat sur ce que signifie précisément donner un nom. Pour le philosophe Gilbert Simondon, l'individuation de l'objet technique, à laquelle participe l'acte de nommer, lui confère une dignité propre[18]. Dès lors, détruire ou jeter cet objet devient moralement problématique. Pour les agents conversationnels, cela poserait la question du jugement que l'on porterait sur l'acte de remettre à zéro la mémoire du chatbot ou d'effacer son historique. L'acte de nommer un objet peut ainsi aboutir à des tensions éthiques en dépit de sa commodité.

L'acte de nommer joue un rôle aussi important que le choix du nom. Cette tendance est tout à fait naturelle : il n'est pas envisageable d'interdire aux utilisateurs de nommer les machines dont ils se servent. En revanche, il importe de prendre conscience de la confusion de statut qui peut être provoquée par cet acte de nommer. Cette confusion peut se révéler néfaste selon les contextes d'usage, qu'il est nécessaire d'examiner au cas par cas. Que le nom soit humain (par exemple « Sophia » ou « Albert ») ou non-humain (par exemple « R2D2 »), son attribution participe d'une dynamique d'anthropomorphisation de la machine et d'individuation de l'agent conversationnel.

Le genre grammatical du nom donné au chatbot, pourvu qu'il soit défini, est un élément significatif. Comme d'autres éléments linguistiques, notamment les pronoms personnels, ce choix peut conduire à l'anthropomorphisme et à des biais de genre. Des limites à la libre créativité de l'utilisateur doivent être définies, notamment pour ce qui concerne l'attribution d'un nom propre genré à un agent conversationnel.

Lorsque l'agent conversationnel lui-même emploie son nom – celui qu'on lui a attribué – dans un dialogue, se pose la question de l'autoréférence : à qui ou quoi exactement renvoie ce nom ? Même si le chatbot est dépourvu de toute corporalité, il inscrit son « identité » virtuelle et influence la perception de la réalité par son interlocuteur. L'emploi du pronom « je » par un agent conversationnel est conceptuellement troublant, mais l'éviter dans tous les contextes, à travers une limitation artificielle du vocabulaire du chatbot, ne serait pas moins problématique.

En effet, le fait qu'un agent conversationnel puisse s'attribuer, à la première personne, un rôle analogue à celui d'un être humain, peut aussi renforcer son utilité : « je suis votre médecin », « je suis là pour vous aider », « je vais vous donner des conseils », etc. Ces rôles, même s'ils ne sont qu'énoncés par le chatbot, se confondent par projection avec les professions et rôles humains. Malgré la puissance de cette projection spontanée, le recours au pronom « je » ne doit aucunement permettre d'attribuer à l'agent conversationnel des connaissances et une responsabilité.

PRÉCONISATION 3

PARAMÉTRER L'IDENTITÉ DES AGENTS CONVERSATIONNELS

Pour éviter les biais, notamment de genre, le choix par défaut des caractéristiques d'un agent conversationnel à usage public (nom, pronoms personnels, voix) doit être effectué de façon équitable à chaque fois que cela est possible. S'agissant des agents conversationnels personnels à usage privé ou domestique, l'utilisateur doit pouvoir modifier ces choix par défaut.

[18] Gilbert Simondon, Du mode d'existence des objets techniques, Paris, Aubier, 1958.

3. MALMENER UN AGENT CONVERSATIONNEL

Les assistants vocaux généralistes se font parfois insulter par les utilisateurs. Toutefois, pour un agent conversationnel, définir ce qu'est une insulte ou la reconnaître dans le flux du dialogue est un problème complexe. On trouve sur internet des exemples de tels défoulements produits par des utilisateurs cherchant à s'amuser, ou par des personnes mécontentes des services reçus. Les services client de nombreux industriels confirment qu'il s'agit d'une pratique répandue. Comme dans la vie de tous les jours, une insulte ne peut être évaluée et jugée que dans un contexte et en fonction des circonstances. S'y ajoutent, dans le cas des agents conversationnels, les stratégies de dialogue du chatbot que définit le concepteur, qui pourraient soit provoquer une insulte (incitation malveillante), soit au contraire tenter de l'éviter.

Le problème du comportement insultant dans un dialogue avec un agent conversationnel est analysé depuis les années 2000[19]. Par exemple, les auteurs de XiaoIce, compagnon virtuel initialement développé par Microsoft Chine en 2014 et disponible en plusieurs langues, reconnaissent que la solution de ce problème a posé un véritable défi de conception[20].

Selon certaines réponses à l'appel à contributions du CNPEN (Annexe 2), insulter ou malmener son ordinateur n'a rien d'immoral. Un chatbot n'étant qu'un programme informatique, dépourvu de compréhension, de conscience et de sensibilité, sa situation ne diffère pas de celle d'une voiture ou d'un réfrigérateur. L'insulter ou le malmener ne serait donc pas un acte moralement répréhensible.

D'autres contributeurs pensent qu'une insulte envers un chatbot constitue un acte moralement dégradant pour celui qui la profère. En effet, en parlant à un agent conversationnel, l'utilisateur ne parle pas avec une autre personne. Il est bien le seul à être conscient de la teneur de ce qui est dit : en quelque sorte, il « reçoit » par effet miroir les invectives qu'il profère. Un argument éthique de « transfert négatif »[21], stipule que cela peut conduire à une modification moralement répréhensible du comportement lorsque l'utilisateur aura acquis une certaine liberté pour répéter de telles phrases devant des interlocuteurs humains.

Cette dernière position s'appuie sur un constat fondamental : l'utilisation du langage ne peut être ni complètement déshumanisée ni entièrement désocialisée. Le simple fait de manier le langage, qui est aussi le moyen de la pensée consciente et du jugement, provoque une projection de traits humains sur la machine. Cette projection ne permet pas de délester la charge morale des mots, en séparant complètement le langage du chatbot des significations, associations et jugements que véhicule le langage humain.

Les insultes adressées aux chatbots permettent de mettre en lumière les limites de l'anthropomorphisation des agents conversationnels en poussant à l'extrême les frontières entre la morale individuelle appartenant à la sphère privée et les mœurs collectives qui se manifestent dans l'espace public. L'utilisateur peut être surpris par le comportement du chatbot, par exemple lorsque celui-ci ne répond à aucune insulte ou lorsqu'une insulte tolérée dans la sphère privée se révèle gênante lorsqu'elle est proférée par un chatbot en public.

PRÉCONISATION 4
TRAITER DES INSULTES

S'il est impossible d'exclure les situations où l'utilisateur profère des insultes envers un agent conversationnel, le fabricant doit les prévoir et définir des stratégies de réponse spécifiques. Notamment, l'agent conversationnel ne devrait pas répondre aux insultes par des insultes et ne pas les rapporter à une autorité. Le fabricant d'un agent conversationnel apprenant doit veiller à exclure de telles phrases du corpus d'apprentissage.

QUESTION DE RECHERCHE 1
RECONNAÎTRE AUTOMATIQUEMENT LES INSULTES

Il est nécessaire de développer des méthodes de caractérisation automatique par les agents conversationnels de propos non désirables, notamment des insultes.

4. MANIPULATION PAR UN AGENT CONVERSATIONNEL

Dans certains cas d'usage, les agents conversationnels sont programmés pour influencer l'utilisateur à travers un choix d'architecture ou de langage. La manipulation par un agent conversationnel peut être directe (y compris par des informations inexactes ou tronquées) ou indirecte, via des stratégies de « nudge ».

LE NUDGE

Le nudge est un terme anglais qui signifie suggestion, incitation ou « coup de pouce ». Il s'agit de pousser doucement la personne dans une direction considérée comme « bonne ». Par exemple, un chatbot pourrait encourager un utilisateur à faire plus de sport en citant l'exemple de ses amis sportifs. Cette théorie d'incitation relativement modérée et non envahissante, qui n'interdit rien et ne restreint pas les options de la personne, a été théorisée par l'économiste Richard Thaler[22].

Sur le plan éthique, il faut distinguer si ce comportement correspond à la recherche d'une finalité que le concepteur pense être bénéfique pour lui-même, pour l'utilisateur ou pour la collectivité. La décision de manipuler ou de tromper l'utilisateur, lorsqu'elle est programmée intentionnellement, doit alors être appréciée en vue de cette finalité. Par exemple, un agent conversationnel pourrait refuser de commander un plat de fast-food car l'utilisateur n'a pas fait assez d'activité physique. Apparaît alors un dilemme : s'il y a contre-indication, faut-il programmer un mensonge (« il n'y en a plus ») ou donner une réponse analytique, soulignant les recommandations du médecin par opposition à la demande de l'utilisateur ?

---

[19] Brahnam, Sheryl. 2005. Strategies for handling customer abuse of ECAS. Abuse: The Darker Side of Human Computer Interaction, pages 62–67.

[20] Li Zhou, Jianfeng Gao, Di Li, Heung-Yeung Shum. The Design and Implementation of Xiaoice, an Empathetic Social Chatbot. arXiv:1812.08989 v2 (2019)

[21] Ph. Brey, The ethics of representation and action in virtual reality", Ethics and Information Technology 1: 5–14, 1999.

[22] R.H. Thaler and C.R. Sunstein. Nudge: Improving Decisions About Health, Wealth, and Happiness. Penguin Books, 2009.

Dans le cas où la manipulation est effectuée par un système de recommandation, le jugement éthique porte sur un équilibre entre le bien-être d'un utilisateur générique, évalué à travers une approche statistique qui s'adresse au plus grand nombre (par exemple, suivre un régime alimentaire équilibré ou faire des exercices physiques), et le bien-être d'une personne particulière, c'est-à-dire de l'utilisateur. La définition d'un tel équilibre engage la responsabilité du concepteur. Si la finalité recherchée est jugée conforme au bien-être du plus grand nombre, cela atténuera le jugement négatif que l'on associe à la manipulation et à la tromperie.

Mais la manipulation reste moralement problématique quelle que soit son utilité. Si le recours à un nudge n'est pas nécessairement moralement condamnable, une tromperie, lorsqu'elle n'est pas présente à la conscience de l'utilisateur, porte atteinte à son autonomie et à sa liberté. À l'échelle de toute la société, la propagation de telles tromperies peut induire, à terme, des manipulations politiques. Cela appelle à définir des limites à la manipulation indépendamment de l'utilité et du contexte d'usage.

MESURES JURIDIQUES RELATIVES À LA MANIPULATION

Le droit de l'Union européenne devrait à terme prévoir des mesures encadrant les manipulations par des systèmes numériques. L'article 5 de la proposition du règlement sur l'intelligence artificielle prévoit d'interdire la mise sur le marché, la mise en service ou l'utilisation de systèmes d'intelligence artificielle qui déploient des techniques subliminales afin d'influer significativement sur le comportement d'une personne d'une manière qui cause à cette personne ou à un tiers un préjudice physique ou psychologique. Le même article interdit les systèmes d'intelligence artificielle qui exploitent la vulnérabilité d'un groupe de personnes en vue d'influer sur le comportement de l'une de ces personnes et de lui causer un dommage. L'article 71 du texte définit les sanctions prévues en cas de méconnaissance de ces interdictions. Par ailleurs, la personne victime d'un préjudice pourrait demander des réparations financières. En outre, le Conseil de l'Europe en appelle à mener « au sein des cadres institutionnels appropriés, des débats publics ouverts, éclairés et inclusifs en vue de donner des orientations sur la limite entre les formes de persuasion admissibles et la manipulation inacceptable » (Déclaration du Comité des ministres du Conseil de l'Europe sur les capacités de manipulation des processus algorithmiques, 13 février 2019).

PRÉCONISATION 5
INFORMER SUR LA MANIPULATION À DESSEIN

Dans l'hypothèse où l'agent conversationnel a été programmé de manière à pouvoir influencer le comportement de l'utilisateur dans le cadre de sa finalité, le fabricant doit informer l'utilisateur de l'existence de cette capacité et recueillir son consentement, qu'il doit pouvoir à tout moment retirer. Le fabricant d'un agent conversationnel influenceur doit permettre aux utilisateurs d'être informés sur la nature, l'origine et les modalités de diffusion des messages qui sont émis par cet agent, et leur demander d'être vigilants avant de propager ces messages.

PRÉCONISATION 6
EVITER LA MANIPULATION MALVEILLANTE

Le fabricant doit veiller à éviter la possibilité technique de manipulations malveillantes ou de menaces proférées par l'agent conversationnel. L'utilisateur doit avoir la capacité de signaler certaines expressions non souhaitées en vue d'une modification de l'agent conversationnel par le concepteur.

INFLUENCEURS VIRTUELS

Des agents conversationnels appelés « influenceurs virtuels » sont de plus en plus présents sur les réseaux sociaux, tels Twitter ou Instagram. Ces influenceurs virtuels imitent les humains et manipulent les utilisateurs, notamment en propageant des mésinformations ou de la désinformation[23].
Parmi les influenceurs virtuels, Lil Miquela, créée en 2016, apparaît sur le réseau social Instagram et a actuellement plus de trois millions d'abonnés. À la fois chatbot et personnage animé, Lil Miquela prend les traits d'une jeune femme se faisant l'égérie de marques reconnues ou s'affichant auprès de célébrités du monde réel. Cette influenceuse virtuelle s'engage contre le racisme, le sexisme et les violences policières, et relate même une « fausse agression » dont elle aurait été victime. Elle joue sur l'empathie et l'ambiguïté de son personnage pour attirer l'intérêt des internautes.

Le cas des mensonges émanant d'un chatbot est particulièrement compliqué. En effet, tout mensonge n'est pas moralement condamnable : d'autres principes, comme la pudeur, la générosité, l'utilité, la justice ou la paix, peuvent pousser les êtres humains à mentir. Le mensonge socialement acceptable, parfois qualifié de « mensonge blanc », qui ne porte pas préjudice à autrui, ou le mensonge par omission en sont des exemples. Confronté à certaines questions (par exemple, « est-ce que j'ai un cancer ? »), le chatbot doit-il refuser de répondre et renvoyer vers un interlocuteur humain ?

MESURES JURIDIQUES RELATIVES AU MENSONGE

La question qui se pose juridiquement n'est pas celle de la responsabilité de l'agent conversationnel en cas de mensonge, dès lors que les systèmes d'intelligence artificielle n'ont pas de personnalité juridique, mais celle de son fabricant. Les textes juridiques sur ce sujet sont assez limités car ils couvrent essentiellement la formation du contrat. L'article 1104 du code civil impose une exigence de bonne foi dans les relations contractuelles. L'article 1112-1 prévoit une obligation d'information pour la partie qui connaît une information dont l'importance est déterminante pour le consentement de l'autre, dès lors que cette dernière ignore cette information ou fait confiance à son cocontractant. En outre, pour que le consentement soit éclairé, il ne doit pas être obtenu par dol[24], sous peine de nullité du contrat (article 1137). Par ailleurs, les pratiques commerciales déloyales visant à tromper le consommateur sont prohibées par le code de la consommation (article L. 120 et suivant). L'abus de faiblesse est sanctionné par le code pénal (article 223-15-2).

[23] Voir bulletin de veille n°2 du CNPEN (https://www.ccne-ethique.fr/fr/actualites/comite-national-pilote-dethique-du-numerique-bulletin-de-veille-ndeg2)

[24] L'article 1137 du Code civil prévoit le dol : « Le dol est le fait pour un contractant d'obtenir le consentement de l'autre par des manœuvres ou des mensonges. Constitue également un dol la dissimulation intentionnelle par l'un des contractants d'une information dont il sait le caractère déterminant pour l'autre partie. »

Un agent conversationnel est rarement doté d'une fonction d'évaluation des énoncés qu'il profère, de leur vérité ou fausseté. Même lorsque cette fonction est présente, il s'agit d'un calcul formel, qui ne permet pas d'évaluer le sens d'un énoncé. Si les données mises à disposition de l'agent conversationnel (une base de données spécifique, les paroles proférées par l'utilisateur ou extraites d'Internet) contiennent des phrases mensongères, le chatbot ne l'identifiera pas facilement. La vérité ne peut donc être pour un chatbot que le résultat d'une évaluation algorithmique.

L'agent conversationnel n'a pas de compréhension « humaine » du sens des énoncés qu'il manipule. En conséquence, lorsqu'un chatbot profère un mensonge, ce processus ne relève d'aucune intention maléfique, ni d'un choix moral. Il s'effectue sans prise de conscience, en procédant par simple réalisation des fonctions programmées avec les données dont dispose le programme. Ces facteurs plaident en faveur de l'absence de jugement moral lorsqu'un chatbot profère un mensonge. En cas de mensonge, la responsabilité du fabricant sera appréciée eu égard aux mesures qu'il a prises pour limiter la manipulation ou, au contraire, qu'il n'a pas prévues dans le code informatique ou lors de la sélection des données.

QUESTION DE RECHERCHE 2

ÉTUDIER LES MENSONGES PROFÉRÉS PAR UN AGENT CONVERSATIONNEL

La portée empirique des mensonges proférés par un agent conversationnel nécessite une étude approfondie. Il est également nécessaire de soustraire l'agent conversationnel à la projection de qualités morales à travers une mise en récit de ses actions explicitement différente de celle qui caractérise les mensonges proférés par les êtres humains.

5. LES AGENTS CONVERSATIONNELS ET LES PERSONNES VULNÉRABLES

Les agents conversationnels sont déployés auprès de personnes vulnérables ou en situation de vulnérabilité[25] dans différents domaines, notamment ceux de la santé et de l'éducation. Les dialogues avec des agents conversationnels laissent des traces sous forme de « logs », qui pourraient contenir des données particulièrement sensibles lorsqu'elles proviennent de personnes vulnérables ou en situation de vulnérabilité. Cette collecte de traces peut être nécessaire pour atteindre les objectifs du système. Il est important d'inclure la collecte, le stockage et l'utilisation de ces traces dans un cadre juridique.

On peut distinguer quelques cas particuliers de l'utilisation des agents conversationnels par les personnes vulnérables ou en situation de vulnérabilité.

Par exemple, les enfants sont naturellement enclins à parler à des objets inanimés comme les jouets[26] ou animaux en peluche. Si ceux-ci peuvent, à leur tour, répondre et interagir, à l'exemple des Furby[27], se crée un lien d'attachement plus fort. À la différence des jouets traditionnels, un jouet intégrant un chatbot peut exercer une influence verbale et émotionnelle sur l'enfant.

PRÉCONISATION 7

ENCADRER L'USAGE DES CHATBOTS DANS LES JOUETS

Dans le domaine ludique, tout particulièrement pour la petite enfance, les autorités publiques se doivent d'évaluer les conséquences des interactions avec des chatbots, susceptibles de modifier le comportement des enfants. Les autorités publiques doivent encadrer l'utilisation des agents conversationnels auprès d'enfants au regard de l'impact de cette interaction sur le développement langagier, émotionnel et culturel de l'enfant.

Dans le domaine de l'éducation, les chatbots peuvent aider les élèves à comprendre des concepts difficiles. Par exemple, les concepteurs des assistants vocaux mettent à disposition des usagers des consignes permettant d'employer leurs systèmes à des fins d'éducation[28]. Cependant, l'apprentissage en interaction avec un chatbot n'est pas équivalent à celui avec un éducateur humain. Par exemple, si un agent conversationnel peut apprendre à un élève une meilleure prononciation d'une langue étrangère en lui indiquant précisément ses erreurs et en l'entraînant sur les répétitions de sons choisis, il peut aussi lui enseigner une langue sensiblement différente de celle qu'on apprend naturellement, avec un vocabulaire limité ou inadapté. Un chatbot pourrait notamment apprendre à son interlocuteur à construire des phrases trop littéraires, sans aucune familiarité stylistique, en appliquant les mêmes stratégies de conversation sans distinction de contexte ou de niveau de conversation. Ou encore, un agent conversationnel pourrait inculquer à un élève une manière de prononcer les sons d'une façon inhumaine, à partir des moyennes statistiques calculées par une machine sur le timbre, l'énergie et le rythme de voix, qui ne ressemblerait pas à celle d'un être humain.

Les agents conversationnels sont souvent utilisés dans l'éducation des enfants autistes ou dans la rééducation des personnes en situation de handicap, grâce à la capacité de la machine de répéter des consignes un grand nombre de fois, ce qui n'est pas toujours le cas de l'éducateur humain[29]. Contrairement à ce dernier, la machine ne « s'impatiente » pas et ne doit pas imiter l'impatience dans l'interaction avec les personnes vulnérables. Sur le plan technique, cela demande des solutions particulières car un apprentissage fondé sur l'imitation du comportement humain, c'est-à-dire l'ensemble de données prélevées sur des éducateurs humains, risque de reprendre aussi ces traits non souhaitables.

QUESTION DE RECHERCHE 3

ÉVALUER LES EFFETS ÉDUCATIFS INÉDITS DES CHATBOTS

Dans le domaine de l'éducation, tout particulièrement pour les enfants vulnérables et pour la petite enfance, les autorités publiques se doivent d'évaluer les conséquences des interactions entre les élèves et le chatbot.

[25] On entend ici par « personnes vulnérables ou en situation de vulnérabilité » les personnes, mineures ou majeures, dont la vulnérabilité est liée à l'âge ou à des déficiences, troubles ou états physiques ou mentaux (par exemple, autisme, maladie d'Alzheimer, phobies, anxiété, dépression, etc.).

[26] https://www.hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2019_05_24_Assistants_vocaux_et_enceintes_connectees_FINAL.pdf

[27] https://www.whoson.com/chatbots-ai/hey-furby-did-the-popular-90s-toy-influence-the-chatbot-timeline/

[28] Voir https://aws.amazon.com/fr/education/alexa-edu/ ou https://dialogs.yandex.ru/store/categories/education_reference

[29] Un exemple d'apprentissage avec le chatbot Siri est décrit déjà en 2014 ; J. Newman, « To Siri with Love », New York Times, 18 Octobre 2014.

Dans le domaine de la santé, les agents conversationnels font partie des outils numériques qui contribuent à répondre aux problèmes récurrents de ce secteur : accès aux soins, pénurie de médecins, déserts médicaux, exécution de tâches répétitives autrement confiées au personnel soignant. L'assistance à la personne grâce à des chatbots pour des motifs de suivis pathologiques, de surveillance ou de conseil médical peut révéler une grande partie de l'intimité du patient ; toutefois, elle s'inscrit nécessairement dans un cadre médical.

L'emploi des chatbots pour le conseil médical est le plus souvent réalisé à travers des applications sur smartphone. Dans ce cas, le chatbot peut donner des conseils de santé directement à l'utilisateur ou collecter les informations de santé afin de les transmettre à un professionnel. Il peut prendre en charge les interrogations des patients qui aspirent à devenir acteurs et responsables de leur santé. Des tâches de soin répétitives sont de plus en plus affectées aux chatbots, par exemple l'information du patient avant et après une opération ou l'éducation et le suivi des patients diabétiques[30]. En outre, parce qu'il s'agit de questions intimes que l'on peut hésiter à partager, il existe des chatbots traitant de sexualité, uniquement accessibles sur smartphone personnel et destinés aux jeunes adultes et adolescents[31].

En psychiatrie, les chatbots permettent notamment de réaliser des entretiens de prévention, de diagnostic et de suivi. Dans ce domaine, les chatbots sont de plus en plus utilisés comme des plateformes de transformation psychique pour faciliter la découverte de soi, de son histoire et de son rapport aux autres. Si, il y a encore peu de temps, ces systèmes réalisaient encore des tâches simples et répétitives sous forme de questionnaires, l'arrivée de chatbots qui imitent le comportement des psychiatres peut être source de nouvelles tensions éthiques. En général, les psychiatres consacrent les premiers entretiens à acquérir la confiance du patient. Or, certaines personnes accordent plus aisément leur confiance en dialoguant avec un chatbot plutôt qu'avec une personne humaine. Cet effet provient du fait que le patient perçoit le chatbot comme neutre : il n'exprime pas de jugement moral et ne provoque pas de sentiment de culpabilité chez le patient, comme cela peut être le cas avec un interlocuteur humain[32]. Cela est d'autant plus vrai que les patients, notamment les plus vulnérables, ont le sentiment d'incarnation d'un rôle de « soignant » dans la voix du chatbot, qui est lié à son degré de personnalisation. Ainsi, des personnes peuvent plus facilement donner des informations à des agents conversationnels ; ces informations peuvent ensuite être utilisées par le médecin. Enfin, ces chatbots sont également disponibles en permanence, y compris la nuit, et peuvent concourir à rassurer le patient.

Pour améliorer le bien-être et l'hygiène de vie des personnes, les agents conversationnels peuvent être utilisés afin d'accroître les capacités d'autorégulation (alimentation, sport), notamment quand ils sont connectés à des outils de recueil d'indices physiologiques, comme les montres mesurant le rythme cardiaque, la température, la conductance de la peau, le niveau d'oxygène, etc. La prise en compte de ces informations par un agent conversationnel, qui les énonce explicitement à la personne, amplifie son engagement ou son désengagement auprès de l'agent conversationnel. Ce rétrocontrôle biologique (biofeedback) des mesures corporelles (Quantified Self[33]), en ajoutant une interprétation linguistique aux données numériques collectées, est cependant susceptible de causer une irritation ou un état de stress. Ces systèmes peuvent ainsi influencer les utilisateurs ou les rendre dépendants. Plus la personne est vulnérable, plus l'effet est susceptible d'être important.

PROTECTION JURIDIQUE DES PERSONNES VULNÉRABLES

L'article 5 de la proposition de règlement européen sur l'intelligence artificielle prévoit d'interdire tout système d'intelligence artificielle qui exploiterait la vulnérabilité d'un groupe de personnes en vue d'influer sur le comportement de l'une de ces personnes et de causer un dommage.

PRÉCONISATION 8

RESPECTER LES PERSONNES VULNÉRABLES

Dans le cas du dialogue entre un agent conversationnel et une personne vulnérable, le fabricant de l'agent conversationnel doit veiller à respecter la dignité et l'autonomie de cette personne. Notamment dans le domaine médical, il est nécessaire, dès l'étape de conception des agents conversationnels, d'éviter la confiance excessive en ces systèmes de la part du patient et de veiller à lever la confusion entre l'agent conversationnel et le médecin qualifié.

PRÉCONISATION 9

ANALYSER LES EFFETS DES AGENTS CONVERSATIONNELS COUPLÉS À DES MESURES PHYSIOLOGIQUES

Dans les cas de couplage des agents conversationnels avec des mesures physiologiques (« Quantified Self »), les concepteurs doivent mener des analyses portant sur les risques de dépendance. Les autorités publiques doivent encadrer l'utilisation de ces systèmes au regard de leur impact sur l'autonomie de la personne.

[30] Klonoff, D. C., & Kerr, D. (2016). Digital Diabetes Communication: There's an App for That. Journal of Diabetes Science and Technology, 10(5), 1003–1005.

[31] https://roo.plannedparenthood.org/onboarding/intro ; J. Brixey, R. Hoegen, W. Lan, J. Rusow, K. Singla, X. Yin, R. Artstein, and A. Leuski, "SHIHbot: A Facebook chatbot for Sexual Health Information on HIV / AIDS",

[32] Proceedings of 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue, August 2017, pp. 370-373.
Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Can J Psychiatry. 2019;64(7):456-464. doi:10.1177/0706743719828977

[33] Le quantified self désigne la pratique de la « mesure de soi » et fait référence à un mouvement né en Californie qui consiste à mieux se connaître en mesurant des données relatives à son corps et à ses activités (https://www.cnil.fr/fr/glossaire).

6. LE TRAVAIL ET LES AGENTS CONVERSATIONNELS

Dans le milieu professionnel, les agents conversationnels sont conçus pour différentes applications dans les entreprises et peuvent être déployés facilement sur des plateformes numériques collaboratives. Les chatbots permettent ainsi de répartir les tâches entre collaborateurs, suivre les avancées d'un projet, rappeler les normes, procédures et buts de l'équipe, permettre de comprendre les rôles, contributions, domaines de compétences des collaborateurs, fixer des rendez-vous, suivre les tâches effectuées ou en cours, collecter les tâches affectées pendant les réunions ou bien encore former des salariés[34].

Ainsi l'introduction d'agents conversationnels dans les équipes peut induire des effets organisationnels qui varient selon les secteurs industriels, notamment la hausse de la charge informationnelle et émotionnelle, la baisse potentielle des interactions directes entre collaborateurs humains, l'émergence d'une médiation désincarnée, le sentiment de cohésion ou d'isolement des travailleurs, les effets sur le moral et l'autonomie des employés ainsi que les problèmes d'égalité et de reconnaissance au mérite au sein des entreprises. Il n'existe actuellement pas de résultats systématiques permettant d'évaluer le bien-fondé de ces craintes.

Lorsqu'un agent conversationnel réalise une tâche en entreprise, des questions se posent concernant son contrôle et la responsabilité des propos qu'il émet. Ces systèmes doivent alors être évalués scientifiquement pour éviter les discriminations sociales. Il convient aussi de vérifier les partis pris et les inégalités éventuelles de genre ou d'âge, engendrées à travers le déploiement des chatbots. Dans le respect des droits des travailleurs, l'entreprise doit rendre transparentes les finalités et les procédures internes impliquant les agents conversationnels.

INTERROGATIONS JURIDIQUES AUTOUR DES SYSTÈMES DE RECRUTEMENT

Les chatbots sont utilisés par des responsables des ressources humaines pour le recrutement ainsi que pour le suivi des carrières et la formation des employés. Cet usage commence à être encadré juridiquement. L'article 6 de la proposition de règlement européen sur l'intelligence artificielle ainsi que son annexe III les incluent parmi les systèmes à risque élevé. Une mise en conformité est alors exigée ex ante : processus de gestion des risques, surveillance, détection et correction des biais, documentation technique, journaux d'événements, information des utilisateurs, surveillance humaine, robustesse, sécurité, exactitude, proportionnalité.

PRÉCONISATION 10

DÉFINIR LES RESPONSABILITÉS POUR L'USAGE DES AGENTS CONVERSATIONNELS DANS LE MILIEU PROFESSIONNEL

Le fabricant doit prévoir des mécanismes de contrôle et d'audit afin de faciliter l'attribution de responsabilités au regard du bon fonctionnement ou du dysfonctionnement de l'agent conversationnel dans le milieu professionnel, notamment étudier leurs effets secondaires ou non-intentionnels.

QUESTION DE RECHERCHE 4

ÉTUDIER LES EFFETS DES AGENTS CONVERSATIONNELS SUR L'ORGANISATION DU TRAVAIL

Les autorités publiques et les acteurs privés doivent soutenir des recherches empiriques sur les effets organisationnels de l'introduction d'agents conversationnels dans les équipes selon les secteurs professionnels.

7. LES AGENTS CONVERSATIONNELS ET LA MÉMOIRE DES MORTS

Avec le développement des chatbots, il est devenu envisageable de réaliser des « jumeaux numériques »[35] conversationnels qui reproduisent la parole ou le comportement langagier des personnes décédées[36]. À la suite d'un apprentissage à partir des données langagières prélevées sur cette personne, un chatbot est capable de dialoguer en l'imitant. Si cette technologie des « deadbots » n'est pas encore bien connue du public, plusieurs entreprises dans le monde travaillent dans ce domaine, par exemple HereAfter AI[37] ou MBC Design Center[38].

[34] Voir le rapport du Groupe de travail sur l'avenir du travail du Partenariat mondial sur l'intelligence artificielle (PMIA, ou GPAI en anglais) https://gpai.ai/fr/projets/avenir-du-travail/pmia-groupe-de-travail-sur-l_avenir-du-travail-novembre-2020.pdf

[35] En général, un « jumeau numérique » est un modèle virtuel qui représente au plus près un objet ou un système réel de manière à pouvoir effectuer des simulations représentatives de son fonctionnement et d'évaluer l'impact de modifications ou d'actions sur lui. Des travaux sont menés sur la réalisation de jumeaux numériques dans plusieurs domaines, y compris la médecine (organes ou corps humain).

[36] La loi « Informatique et libertés » intègre dans ses articles 84 à 86 les dispositions de la loi « Pour une république numérique » du 7 octobre 2016 traitant du sort des données personnelles qui sont sur internet après le décès des personnes concernées (notion de mort numérique). Ces dispositions ne sont pas spécifiques à la technologie des agents conversationnels et n'intègrent pas la possibilité de créer des « jumeaux numériques » après la mort. En revanche, l'article 85 dispose que : « Toute personne peut définir des directives relatives à la conservation, à l'effacement et à la communication de ses données à caractère personnel après son décès. Ces directives sont générales ou particulières. »

[37] https://www.hereafter.ai

[38] https://www.ubergizmo.com/2021/01/virtual-reality-husband-meet-deceased-wife/

Typiquement, les agents conversationnels ne répètent pas simplement les données d'apprentissage, mais possèdent la capacité à générer des propos nouveaux, que la personne imitée n'a jamais proférés de son vivant. C'est le cas de l'invention de la société Microsoft, qui a déposé un brevet combinant un deadbot et un agent conversationnel de type transformeur[39]. Qu'ils soient purement conversationnels ou dotés de la fonction d'imitation visuelle, certains de ces systèmes peuvent avoir un dialogue très vraisemblable, favorisé éventuellement par la capacité de simulation d'émotions du chatbot. L'interlocuteur humain peut avoir réellement l'impression de se trouver en présence de la personne ainsi imitée, même s'il est explicitement informé qu'il s'agit d'une machine. Dans un exemple saisissant, un jeune homme canadien a utilisé les capacités du réseau de neurones de type transformeur GPT-3 afin d'imiter le dialogue avec sa petite amie décédée[40]. Cette thématique suscite depuis quelques années des réactions passionnées des utilisateurs des « deadbots » et du public[41].

UN EXEMPLE DE « DEADBOT » AU JAPON

En 2018, un projet japonais[44] propose aux proches d'une personne décédée d'utiliser un robot humanoïde équipé d'un masque imprimé en 3D représentant le visage de cette personne. Les utilisateurs interagissent avec ce robot grâce à un système qui permet d'imiter certains traits de la personnalité du défunt en se servant de sa parole et de ses gestes préenregistrés, sans toutefois innover ou produire des contenus que la personne n'avait pas réellement proférés de son vivant. La durée de vie de cette machine, par construction, n'est que de 49 jours, ce qui correspond à la durée traditionnelle de la période de deuil dans la culture japonaise.

Certains voient dans cette technologie, qu'ils trouvent fascinante, une opportunité pour « dépasser » ou « tromper » la mort (Annexe 2). D'autres effrayés par cette extension illusoire de la vie, pensent que cela porte atteinte au principe de respect de la dignité de la personne humaine, même si cette notion morale majeure[42] est difficile à cerner[43]. D'après eux, la génération de propos nouveaux, émanant faussement d'une personne décédée, ne devrait pas être autorisée après sa mort et toute ingérence numérique avec cet élément fondamental de la nature humaine devrait être interdite. Le sujet des deadbots cristallise donc les fantasmes et les inquiétudes, et questionne notre conception de la dignité humaine.

Les conceptions de la mort et de ses différentes étapes varient avec les cultures et les époques. Les rites funéraires prennent des formes très différentes suivant les cultures (momification, crémation, inhumation, etc.) et peuvent s'étendre sur plusieurs mois. De même la relation posthume aux corps et aux esprits des morts varie suivant les religions et les cultures. Elle peut se traduire par un véritable culte des personnes disparues. La littérature occidentale depuis Homère et Virgile, sans oublier Dante et Molière, contient également de nombreux exemples de dialogues avec les morts. Dans la culture japonaise marquée par diverses traditions religieuses, en particulier par le shintoïsme, les fantômes ou les sosies des morts apparaissent abondamment dans la littérature ou la production cinématographique.

Les photographies et les enregistrements audio et vidéo permettent déjà d'assurer une capacité de remémoration d'une personne après sa disparition. Les chatbots pourraient n'être qu'une étape supplémentaire sur cette voie ouverte par la technologie depuis l'invention de l'écriture. La capacité des agents conversationnels à générer des propos nouveaux, que la personne imitée n'a jamais proférés de son vivant, nécessite dès lors une attention particulière car elle les différencie des techniques de remémoration plus anciennes.

Des propos inédits et pourtant vraisemblables ne peuvent être attribués à la personne imitée que sur un mode conditionnel : cette personne aurait pu dire de telles phrases, même si elle ne les avait pas proférées réellement. Leur effet est cependant bien réel. La réaction du jeune homme canadien ayant entrainé un chatbot avec les données langagières de sa petite amie décédée en offre un exemple troublant : « Le plus mystérieux de tout : le chatbot semblait percevoir les émotions. Il savait comment dire une phrase juste, au bon moment et en mettant le bon accent »[45]. La responsabilité pour ces propos ressemblants mais inventés, est un problème éthique et juridique inédit. Il doit être analysé dans le cadre du débat actuel sur l'intelligence artificielle en France et en Europe. Au demeurant, la société OpenAI, propriétaire de GPT-3, a décidé de fermer l'accès à ce réseau de neurones à l'informaticien californien ayant contribué à développer le chatbot utilisé par le jeune homme canadien pour créer le double de son amie décédée[46].

[39] https://www.forbes.com/sites/barrycollins/2021/01/04/microsoft-could-bring-you-back-from-the-dead-as-a-chat-bot/

[40] https://www.sfchronicle.com/projects/2021/jessica-simulation-artificial-intelligence/

[41] https://www.liberation.fr/futurs/2017/07/19/un-journaliste-discute-avec-son-pere-decede-grace-a-un-programme-qu-il-a-cree_1584849

[42] En philosophie morale, le concept de dignité humaine est à l'origine de ce qu'Emmanuel Kant définit comme un impératif catégorique ; elle se retrouve à l'article premier de la Déclaration universelle des droits humains de 1948 : « Tous les êtres humains naissent libres et égaux en dignité et en droits ».

[43] La notion de dignité humaine est juridiquement difficile à cerner. Elle n'est pas définie par un texte normatif et découle seulement, en droit interne, de la jurisprudence du Conseil constitutionnel : il s'agit d'un principe à valeur constitutionnelle défini pour la première fois dans la décision n° 94-343/344 du 27 juillet 1994, dite bioéthique. Il a été ensuite appliqué en droit pénal, en cas de privation de liberté et d'hospitalisation sans consentement. Le Conseil d'État, dans son arrêt du 27 octobre 1995, Commune de Morsang-sur-Orge, n° 136727, en fait une composante de l'ordre public : l'attraction du « lancer de nain » consistant à faire lancer une personne de petite taille par des spectateurs, ce qui conduit à utiliser comme un projectile une personne affectée d'un handicap physique, porte atteinte à la dignité de la personne humaine et, en conséquence, l'autorité investie du pouvoir de police municipale pouvait l'interdire, alors même que des mesures de protection avaient été prises pour assurer la sécurité de la personne concernée qui avait donné son consentement et était rémunérée pour cette prestation.

[44] https://starts-prize.aec.at/en/digital-shaman-project/

[45] https://www.sfchronicle.com/projects/2021/jessica-simulation-artificial-intelligence/

[46] https://gadgets.ndtv.com/internet/news/openai-chatbot-gpt-3-samantha-shut-down-dilute-jason-rohrer-possible-misuse-2537388

Le débat sur la technologie des « jumeaux numériques » des personnes humaines n'en est qu'à ses débuts. Le CNPEN attire l'attention des parties prenantes sur la complexité éthique du problème. La création de « jumeaux numériques » des personnes, notamment des défunts, doit être soumise à une interrogation éthique concernant d'abord la démarche elle-même, la raison pour laquelle un tel système serait réalisé. Elle doit se poursuivre de manière continue si le projet est réalisé, à l'étape de conception, de collecte de données avec le consentement de la personne dont elles émanent, ainsi qu'à l'étape d'utilisation. Il est également nécessaire de définir des limites juridiques à cette technologie, en les faisant au besoin évoluer à l'aune des avancées futures. Le consentement pour le recueil des données du défunt doit être encadré selon sa temporalité (avant la mort / après le décès) et son auteur (la personne de son vivant / ses héritiers, ce qui pose un problème compte tenu de la non transmissibilité des données personnelles selon le droit en vigueur). La personne décédée ne pouvant retirer son consentement, la capacité d'usage du « deadbot » est donc entièrement laissée à l'utilisateur, à moins d'un cadre légal spécifique. Un autre risque issu des technologies de « jumeaux numériques », est celui d'usurpation, à travers un agent conversationnel, de l'identité d'une personne vivante ou décédée[47].

L'agent conversationnel qui imite une personne décédée est souvent utilisé par quelqu'un qui a connu la personne de son vivant. Même si l'utilisateur sait que l'échange n'émane que d'un chatbot, et non de la personne qu'il avait connue, il n'en a pas en permanence conscience. L'utilisateur parvient alors à réaliser son envie de se remémorer du défunt en se laissant parfois entraîner dans l'illusion de sa présence. Cette situation se rapprocherait des séances de spiritisme dans les pratiques antérieures, à la différence que l'utilisateur sait qu'il interagit avec une machine.

Dans certains cas, il pourrait en résulter une altération du jugement, problématique tant sur le plan cognitif que moral. Notons que le chatbot n'a pas de compréhension de la signification du langage ni du contexte. Ses propos pourraient ainsi provoquer un effet de « vallée de l'étrange » pour l'interlocuteur : soit le chatbot profère des propos offensants, soit, après une séquence de répliques familières, il débite une phrase totalement différente de ce qu'aurait pu dire la personne imitée. Dans ces circonstances, l'utilisateur se demanderait naturellement s'il s'agit d'une erreur, d'un contresens ou d'un propos signifiant. Il pourrait subir un changement psychologique rapide et douloureux. Ceci appelle à définir des limites des chatbots qui simulent la parole des personnes décédées ou sont présentés comme leurs « jumeaux numériques ».

Le respect de la mémoire et de la dignité des morts est un principe largement partagé par les êtres humains. La réalisation d'agents conversationnels simulant les personnes décédées polarisent déjà les opinions. Une société peut décider d'interdire un tel développement ; mais elle peut également en limiter la réalisation par des mesures d'ordre juridique. Dans ce cas, un cadre spécifique doit être élaboré, ainsi qu'un ensemble de contraintes techniques limitant les effets secondaires, notamment l'éventualité d'effets négatifs sur le processus du deuil.

PRÉCONISATION 11

MENER UNE RÉFLEXION SOCIÉTALE AVANT TOUTE RÉGLEMENTATION DES « DEADBOTS »

À la suite d'une réflexion éthique approfondie à l'échelle de toute la société, le législateur devrait adopter une réglementation spécifique concernant les agents conversationnels qui imitent la parole des personnes décédées.

PRÉCONISATION 12

ENCADRER TECHNIQUEMENT LES « DEADBOTS »

Les concepteurs de « deadbots » doivent respecter la dignité de la personne humaine qui ne s'éteint pas avec la mort, tout en veillant à préserver la santé mentale des utilisateurs de tels agents conversationnels. Des règles doivent être définies concernant notamment le consentement de la personne décédée, le recueil et à la réutilisation de ses données, le temps de fonctionnement d'un tel chatbot, le lexique utilisé, le nom qui lui est attribué ou encore les conditions particulières de son utilisation.

8. EFFETS À LONG TERME DES AGENTS CONVERSATIONNELS

En plus des vulnérabilités avérées discutées au paragraphe II.5, il existe des risques émergents pour tous les utilisateurs des agents conversationnels, liés à de nombreux effets potentiels comme le manque d'interaction avec autrui, les biais cognitifs ou la crédulité. Ces risques peuvent résulter de l'interaction avec les chatbots et de la confiance – parfois excessive – que les utilisateurs accordent aux agents conversationnels selon différents rôles qu'endossent ces derniers (professeur, banquier, médecin ou ami). De plus, l'évolution des normes de comportement induite par les agents conversationnels est susceptible de créer de nouvelles vulnérabilités individuelles et collectives. Cette évolution est déjà en cours, lancée via l'interaction des utilisateurs avec des agents conversationnels omniprésents, embarqués sur les smartphones (SIRI ou Google Assistant) ou sur des enceintes vocales (Alexa Amazon, Google Home).

[47] En précisant notamment l'article 226-1-4 du Code pénal : « le fait d'usurper l'identité d'un tiers ou de faire usage d'une ou plusieurs données de toute nature à permettre de l'identifier en vue de troubler sa tranquillité ou celle d'autrui, ou de porter atteinte à son honneur ou à sa réputation est puni d'un an d'emprisonnement et de 15 000 € d'amende. Cette infraction est punie des mêmes peines lorsqu'elle est commise sur un réseau de communication au public en ligne ».

À moyen et long termes, l'utilisation des chatbots avec un effet d'habituation peut avoir une incidence durable sur le langage humain et sur l'évolution des normes de comportement. Par exemple, si les chatbots répondent par des phrases courtes, linguistiquement pauvres, sans aucune politesse, les personnes risquent d'imiter ces caractéristiques langagières lorsqu'elles s'adressent à d'autres personnes. Ces effets sont encore incertains et il est nécessaire de les étudier de manière prospective en mesurant l'impact durable sur les utilisateurs. Les interactions avec les chatbots sont susceptibles d'influer sur les modes de vie, les opinions et les décisions des êtres humains. Il importe de prendre conscience à tous les niveaux, de l'ingénieur au politique, de l'ampleur et de l'étendue des effets futurs des agents conversationnels sur les croyances, opinions et décisions des utilisateurs, notamment des effets de masse. Les performances des modèles du langage utilisés dans les systèmes d'apprentissage machine les plus récents (par exemple, les réseaux de neurones comme GPT-3[48] ou LamDA[49]) marquent un vrai tournant dans le développement de ces technologies. Aujourd'hui, les réseaux de neurones de type transformeur permettent de choisir la stratégie de dialogue et de générer des réponses. Ils dépassent de loin les capacités des enceintes connectées de générations antérieures, étudiées dans la littérature des sciences humaines et sociales[50]. Ces modèles intègrent de très grands volumes de données collectées sur le web, les recoupent et les retranscrivent, de manière souvent non-reproductible, selon la demande, sans pour autant en abstraire le sens ou raisonner comme un être humain.

La machine ne fait que des calculs : elle donne une réponse calculée lorsqu'elle se trouve face à une question. Pour autant, cela n'empêche pas les réseaux de neurones de type transformeur de produire des phrases que l'interlocuteur humain trouvera originales, qui ne reproduisent aucune de celles utilisées lors de l'apprentissage. Les utilisateurs projettent alors des significations sur ces propos originaux. Cette projection complexifie l'attribution de responsabilité sur les propos tenus par la machine. Elle montre qu'il ne faut pas considérer cette technologie comme « neutre » car, malgré son caractère asémantique, elle participe de la construction du sens éthique et politique des énoncés.

En mémorisant nos paroles et actions, les agents conversationnels sont capables d'en déduire des informations sur nos opinions, décisions ou encore nos points de faiblesse. Par exemple, un chatbot est capable de rappeler des souvenirs que l'utilisateur avait oubliés. Un autre effet peut être celui d'une incitation à se dévoiler davantage. À long terme, la notion même d'intimité d'une personne peut évoluer sous l'influence des agents conversationnels.

GATEBOX

La société japonaise Gatebox[51] commercialise Azuma Hikari, un chatbot représentant une « petite amie virtuelle » en hologramme, capable d'allumer et d'éteindre la lumière, d'envoyer des SMS, de reconnaître les gens et de discuter avec eux. Cet attachement provoque une vulnérabilité émergente due à la dépendance affective et à la relation intime qui s'installe au cours du temps.

Un chatbot « ange gardien » est, par exemple, un agent conversationnel qui a pour fonction la protection permanente des données personnelles de l'utilisateur. Il veille ainsi au respect de la vie privée de son interlocuteur.

Des interactions quotidiennes avec un agent conversationnel de ce type ou avec un « ami virtuel » modifient la notion de vie intime et le rapport aux autres. Elles peuvent créer des dépendances notamment pour les enfants dont le développement est fondé sur une relation privilégiée avec leur entourage auquel participent désormais les chatbots. Dans le même temps, les chatbots peuvent pallier les divers manques ou répondre à des traumatismes subis. Cette fonction de l'agent conversationnel correspond au besoin de la personne d'être rassurée et de recevoir des réponses à ses questions. Le chatbot apparaît alors comme un modèle ou un miroir éducatif des comportements souhaités. À l'échelle de la société, ces effets peuvent produire à long terme une évolution sensible de la condition humaine. La co-adaptation langagière entre les utilisateurs humains et les agents conversationnels est le moteur de cette transformation.

PRÉCONISATION 13

ENCADRER LE DÉPLOIEMENT DES CHATBOTS « ANGE GARDIEN »

Les autorités publiques doivent encadrer l'utilisation des agents conversationnels de type « ange gardien », qui sont capables de protéger les données d'une personne, afin de limiter le paternalisme et respecter l'autonomie humaine.

QUESTION DE RECHERCHE 5

ÉTUDIER LES EFFETS DE L'UTILISATION DES CHATBOTS À LONG TERME

Les autorités publiques et les acteurs privés doivent investir dans des recherches sur les effets à long terme et les conséquences de l'utilisation des agents conversationnels sur l'être humain et la société. Tous les acteurs de la société doivent rester vigilants quant aux effets futurs des agents conversationnels sur les croyances, opinions et décisions des utilisateurs, notamment des effets de masse, et éviter de considérer cette technologie comme neutre ou dépourvue de signification éthique et politique.

Le marché des agents conversationnels est en pleine expansion, notamment grâce aux réseaux de neurones de type transformeur (voir chapitre III). Du fait du recours aux technologies d'apprentissage à partir de très grandes bases de données et à celles d'adaptation en continu à l'utilisateur, les agents conversationnels seront de plus en plus consommateurs de puissance de calcul et en taille de la mémoire. Le déploiement accéléré des chatbots pose donc la question de la dépense d'énergie, même si elle n'est pas spécifique aux agents conversationnels.

QUESTION DE RECHERCHE 6

ÉTUDIER L'IMPACT ENVIRONNEMENTAL

Les autorités publiques et les acteurs privés doivent mener des études sur l'impact énergétique et environnemental de la technologie des agents conversationnels.

[48] https://arxiv.org/abs/2005.14165

[49] https://blog.google/technology/ai/lamda/

[50] Par exemple, Usage et valeur 62 (10/2019) ; Réseaux 2020/2-3 (N° 220-221) ; H. Kempt, Chatbots and the Domestication of AI: A Relational Approach, Springer, 2020.

[51] https://www.gatebox.ai

III. PRINCIPES ÉTHIQUES DE CONCEPTION DES AGENTS CONVERSATIONNELS

La technologie des agents conversationnels soulève des interrogations éthiques liées à la conception de ces systèmes. Certaines questions se posent pour tous les chatbots, même ceux qui suivent un algorithme déterministe avec un choix de réponses prédéfinies en nombre limité. D'autres sont spécifiques aux chatbots qui traitent des émotions et plus largement du comportement des utilisateurs, à ceux qui utilisent un apprentissage adaptatif, ou encore à ceux qui utilisent les réseaux de neurones de type transformeur pour la gestion du dialogue. Les questions éthiques majeures de conception des chatbots, ont été également soulevées dans la consultation publique (voir annexe 2) et sont regroupées sous les cinq sections suivantes.

1. ÉTHIQUE PAR CONCEPTION

La notion d'« ethics by design » (« éthique par conception »)[52] repose sur l'idée de respect des valeurs fondamentales lors de la conception d'un système technique. Elle est définie au sein de cadres théoriques et méthodologiques comme la conception sensible aux valeurs[53], le « value-sensitive design »[54] ou le « Technology Assessment »[55]. Ces approches, qui sont en développement depuis plus de trois décennies, visent à intégrer de manières différentes les valeurs humaines dans les processus de conception des systèmes techniques. Cela ne signifie pas pour autant que les valeurs sont directement traduites dans le code informatique ; leur intégration exige un procédé de conception impliquant les programmeurs, les entrepreneurs, les utilisateurs et les décideurs politiques. Ces approches fournissent ainsi une trame pour analyser la redistribution des responsabilités qu'induit la diffusion des systèmes d'intelligence artificielle, notamment des agents conversationnels. Elles fournissent également une trame pour la formation et l'éducation.

La démarche d'évaluation, concept étymologiquement lié à la notion de valeur, fait partie intégrante des approches d' « éthique par conception ». Si le cadre éthique est fixé en termes de valeurs, il s'agit d'évaluer le degré de correspondance entre ces valeurs et le fonctionnement d'un système. L'exemple le plus évident est celui de l'évaluation des biais induits dans la construction et l'entraînement des systèmes algorithmiques reposant sur l'apprentissage statistique à partir de grands corpus de données. La non-discrimination des groupes d'utilisateurs par un système d'intelligence artificielle ne doit pas seulement être proclamée, mais bien mesurée à l'aide d'indicateurs quantitatifs spécifiques. Il existe un corpus de travaux scientifiques sur l'évaluation des biais, y compris pour les agents conversationnels, qui participent de la démarche « éthique par conception » de ces systèmes. Plusieurs acteurs industriels, y compris les géants du numérique[56], intègrent déjà dans leur procédé de conception des outils de mesure des biais explicites ou implicites, contenus dans leurs produits.

PRINCIPE DE CONCEPTION 1

« ÉTHIQUE PAR CONCEPTION » DES AGENTS CONVERSATIONNELS

Les concepteurs d'un agent conversationnel doivent analyser, en phase de conception, chacun des choix technologiques susceptibles de provoquer des tensions éthiques. Si une tension potentielle est identifiée, ils doivent envisager une solution technique visant à diminuer ou à faire disparaître la tension éthique, puis évaluer cette solution dans des contextes d'usage réalistes.

QUESTION DE RECHERCHE 7

DÉVELOPPER LES MÉTHODOLOGIES « ÉTHIQUE PAR CONCEPTION » POUR LES CHATBOTS

Les autorités publiques doivent soutenir des recherches afin d'élaborer des méthodologies « éthique par conception » adaptées au développement des agents conversationnels.

2. BIAIS ET NON-DISCRIMINATION

Les phrases produites par un agent conversationnel peuvent contenir des biais : par exemple, un corpus de paroles enregistrées peut contenir uniquement des voix d'adultes alors que le système est censé interagir aussi avec des enfants, ou un corpus de textes peut utiliser statistiquement plus fréquemment des pronoms de genre féminin plutôt que masculin. Si les algorithmes peuvent être utilisés de manière positive pour révéler ces biais, ils intègrent aussi des biais de nature sociale ou historique. Le système reproduira alors ces biais, sauf s'il est équipé de modules spécialement conçus dans le but de les corriger, ce qui présuppose déjà la connaissance des biais possibles et la capacité à les corriger. Or, certains biais pourraient ne pas être connus à l'avance. La présence des biais dans le comportement des agents conversationnels est une source majeure de conflits éthiques ou de discriminations directes : une personne pourrait être traitée de manière moins favorable qu'une autre au regard de critères tels que, notamment, l'âge, le sexe, le genre, le handicap ou la couleur de la peau, pour l'accès à un emploi, à un logement ou à un droit[57]. Ils peuvent aussi entraîner des discriminations indirectes : par exemple, les premières personnes à passer un entretien d'embauche pourraient être désavantagées si les paramètres du chatbot qui analyse ces entretiens évoluent à la suite d'un apprentissage adaptatif, sous l'influence des données des candidats déjà auditionnés.

---

[52] Voir les livrables des projets Horizon-2020 SIENNA et Sherpa financés par la Commission européenne.

[53] J. van den Hoven et al. (eds.), Handbook of Ethics, Values, and Technological Design. Springer, 2015.

[54] B. Friedman and D.G. Henry. Value Sensitive Design. Shaping Technology with Moral Imagination. MIT Press, 2019.

[55] A. Grunwald et R. Hillerbrand. Handbuch Technikethik. 2e éd. J.B.Metzler, 2021.

[56] R. K. E. Bellamy et al., «AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias,» IBM Journal of Research and Development, vol. 63, no. 4/5, pp. 4:1-4:15, 1 July-Sept. 2019, doi: 10.1147/JRD.2019.2942287.

[57] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623).

PRINCIPE DE CONCEPTION 2
RÉDUIRE LES BIAIS DANS LE LANGAGE

Pour réduire les biais dans le langage et tenter d'éviter les effets de discrimination, notamment culturels, les concepteurs doivent mettre en œuvre une solution technique à trois niveaux : dans la mise en œuvre de l'algorithme, dans la sélection des paramètres d'optimisation et dans le choix des données d'apprentissage et de validation pour les différents modules des agents conversationnels.

TRAITEMENT JURIDIQUE ET TECHNIQUE DES BIAIS ET DE LA DISCRIMINATION

En 2019 et en 2020, le Défenseur des droits et la CNIL ont mis l'accent sur les risques de discrimination pouvant résulter des biais algorithmiques[58]. Une réflexion est également engagée au niveau européen pour adapter le cadre réglementaire afin d'appréhender de tels risques. Ainsi le Conseil de l'Europe, dans ses nouvelles lignes directrices de la convention 108, a, en 2020, préconisé que les développeurs, fabricants et prestataires de service devaient éviter tout biais potentiel, y compris les biais non intentionnels ou cachés, ainsi que les risques de discrimination. Le Parlement européen, dans sa résolution du 19 février 2019, a indiqué « que les résultats doivent être révisés pour éviter toute forme de stéréotype, de discrimination ou de biais et que l'IA doit au besoin être employée pour repérer et corriger tous les biais humains ». La proposition de Règlement européen de l'intelligence artificielle, publiée par la Commission européenne le 21 avril 2021, contient des mesures destinées à limiter les biais discriminatoires et met la notion de supervision humaine au cœur de la lutte contre ces biais. Il est notamment prévu que les ensembles de données d'entraînement, de validation et de test doivent être soumis à des pratiques appropriées de gouvernance et de gestion des données, afin de prendre en considération d'éventuels biais. Il n'est pas précisé comment les systèmes seront testés pour détecter ces biais : faut-il prendre comme référence l'égalité des chances ou l'égalité des résultats ou d'autres critères ?

3. TRANSPARENCE, REPRODUCTIBILITÉ, INTERPRÉTABILITÉ ET EXPLICABILITÉ

Les principes de transparence, reproductibilité, interprétabilité et explicabilité sont essentiels dans le cadre de l' « éthique par conception », même s'il existe dans chaque situation concrète des tensions entre ces principes. Leur réalisation dépend du contexte et doit être pensée à l'égard du principe de proportionnalité et dans le respect des droits fondamentaux.

La transparence d'un système implique, entre autres, que son fonctionnement ne soit pas perçu comme opaque ou incompréhensible par l'utilisateur. Dans le cas d'un agent conversationnel, elle s'appuie particulièrement sur la traçabilité des répliques que celui-ci a produites.

PRINCIPE DE CONCEPTION 3
ÉNONCER LES FINALITÉS DU CHATBOT

Le concepteur doit veiller à ce qu'un agent conversationnel communique sa finalité à l'utilisateur de manière claire et aisément compréhensible au moment adéquat, par exemple au début ou à la fin de chaque conversation.

PRINCIPE DE CONCEPTION 4
TRANSPARENCE ET TRAÇABILITÉ DU CHATBOT

Dans le respect du RGPD, l'agent conversationnel devrait pouvoir sauvegarder une partie du dialogue (dont l'étendue reste à définir) à des fins de preuve en cas de litige ou encore pour satisfaire à des contraintes de sécurité. Ce besoin introduit une tension avec la protection des données à caractère personnel. L'architecture des chatbots, les connaissances utilisées et les stratégies de dialogue doivent être rendues accessibles pour l'audit ou le traitement d'éventuels problèmes juridiques. Cette préconisation pourrait déboucher sur une mesure d'ordre réglementaire définissant les conditions précises de son application.

PRINCIPE DE CONCEPTION 5
TRAITEMENT DES DONNÉES COLLECTÉES PAR LES AGENTS CONVERSATIONNELS

À l'image de l'encadrement existant des données de santé, il est nécessaire d'élaborer des règles éthiques et juridiques pour la collecte, le stockage et l'utilisation des traces linguistiques des interactions entre les agents conversationnels dans le respect du RGPD.

Il est aujourd'hui possible de personnaliser fortement les répliques d'un chatbot. Par exemple, l'historique des échanges entre l'agent conversationnel et le patient, incluant des informations sur le psychisme ou les croyances de celui-ci, peut servir à améliorer son traitement. Le profilage ou l'analyse comportementale pendant le dialogue est une méthode permettant d'estimer le profil psychologique, économique ou autre, de l'utilisateur.

MESURES JURIDIQUES RELATIVES AU PROFILAGE

L'article 4 du RGPD définit le profilage comme toute forme de traitement automatisé de données à caractère personnel consistant à utiliser ces données pour évaluer certains aspects d'une personne, notamment pour analyser ou prédire des éléments concernant le rendement au travail, la situation économique, le comportement, etc. Les décisions qui découlent d'un profilage sont encadrées par l'article 47 de la loi informatique et libertés et l'article 22 du RGPD, dès lors qu'elles sont susceptibles d'avoir des effets sur l'individu. Selon l'article 22 du RGPD, « la personne concernée a le droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques la concernant ou l'affectant de manière significative de façon similaire. » Trois exceptions sont prévues à cette garantie : le consentement explicite de la personne, l'existence d'un contrat rendant nécessaire l'existence d'une prise de décision automatisée, des dispositions légales spécifiques autorisant une décision automatisée. Par ailleurs, en vertu des articles 13- 2 f) et 14- 2 g) du RGPD, les personnes qui font l'objet d'une décision entièrement automatisée doivent être informées, lors de la collecte de leurs données et « à tout moment sur leur demande, de l'existence d'une telle décision et de sa logique sous-jacente ».

[58] https://www.defenseurdesdroits.fr/fr/communique-de-presse/2020/05/algorithmes-et-discriminations-le-defenseur-des-droits-avec-la-cnil

PRINCIPE DE CONCEPTION 6

INFORMER SUR LES CAPACITÉS DES AGENTS CONVERSATIONNELS

Dans un souci de transparence, l'utilisateur doit être informé de manière adaptée, claire et compréhensible, oralement ou par écrit, de la collecte des données, des capacités de l'agent conversationnel liées à l'adaptation aux données qu'il collecte en cours d'utilisation et au profilage.

Les agents conversationnels utilisant l'apprentissage statistique posent aussi un problème de reproductibilité : un modèle obtenu par apprentissage peut donner à chaque instance des résultats proches dans son espace de calcul, mais éventuellement différents du point de vue de l'utilisateur, bien que les paramètres du système et les données d'entrée soient les mêmes.

QUESTION DE RECHERCHE 7

REPRODUCTIBILITÉ DES PROPOS DES AGENTS CONVERSATIONNELS

La reproductibilité nécessite de mémoriser les données, mais aussi de définir la bonne mesure de répétition dans les propos du chatbot. Ces questions doivent être étudiées.

Si les acteurs du numérique s'accordent sur l'importance des principes d'explicabilité et d'interprétabilité, leur contenu semble toutefois varier selon les sources. En général, ils traduisent la mesure dans laquelle un observateur peut comprendre la décision de la machine et ses causes. Si toute solution technologique est mise en œuvre par son concepteur, cela n'implique pas que celui-ci participe intentionnellement et consciemment à toute prise de décision de l'algorithme. Dans le cas des systèmes d'apprentissage machine, la chaîne causale menant à un résultat est opaque par construction. Les recherches dans le domaine d'IA explicable (xAI) visent à établir des heuristiques d'explication pour ces systèmes.

De manière pragmatique, l'explicabilité d'un agent conversationnel présuppose la mise en œuvre de solutions techniques permettant à l'utilisateur de mieux interpréter le comportement du chatbot en reconstruisant un « raisonnement » cohérent derrière ses répliques, même si en réalité l'agent conversationnel ne raisonne pas. Un chatbot ne « comprend » pas le sens des phrases qu'il génère ou qu'il perçoit. Il n'a pas de sens commun. Il est ainsi susceptible de formuler des phrases qui ne correspondent à aucune réalité humaine (« lait lyrique »), de répondre sans tenir compte du contexte (« Comment vas-tu ? » - « Il fait beau ») ou d'employer un lexique inapproprié. Les effets immédiats sur l'utilisateur provoqués par un tel dialogue peuvent être importants : réaction émotionnelle forte, rupture dans la compréhension, abandon du dialogue ou débranchement du système. Tous ces effets interrogent la responsabilité des concepteurs.

PRINCIPE DE CONCEPTION 7

FAVORISER L'EXPLICABILITÉ DU COMPORTEMENT DU CHATBOT

Les concepteurs doivent développer des solutions favorisant la compréhension par les utilisateurs du comportement des agents conversationnels.

4. INTERACTION AFFECTIVE AVEC L'ÊTRE HUMAIN ET ADAPTATION AUTOMATIQUE

Certains agents conversationnels intègrent un module de prédiction du comportement émotionnel, attentionnel ou intentionnel des êtres humains. Ils peuvent également simuler une expression affective dans leurs répliques écrites ou synthétisées. Enfin, ces informations peuvent être utilisées par le système de dialogue pour choisir une stratégie de réponse ; elles peuvent aussi être intégrées dans les réseaux de neurones artificiels, notamment dans les transformeurs. De nombreux exemples ont été cités dans les sections II.7 et II.8, comme les agents conversationnels qui imitent une petite amie virtuelle (Gatebox) ou une personne décédée (« deadbot »).

Les tentatives pour mettre au point de telles technologies, appelées informatique affective, remontent aux travaux de Rosalind Picard, chercheur du Massachusetts Institute of Technology, qui publie en 1995 un article posant les fondements de cette nouvelle discipline. L'informatique affective regroupe donc trois technologies : la détection des émotions, la prise en compte des états émotionnels dans les stratégies de dialogue et la génération et la synthèse émotionnelles.

L'informatique affective s'appuie sur la notion de théorie de l'esprit, qui désigne l'aptitude cognitive d'un individu à attribuer des états mentaux inobservables (intention, désir, croyance, émotion) à soi-même ou à d'autres individus. Lorsqu'un être humain perçoit une émotion chez autrui, il s'agit de perception subjective qui lui est propre. La perception de la machine est, quant à elle, une estimation probabiliste obtenue par un calcul dans un réseau stochastique, construit à partir de données émotionnelles produites par de nombreux sujets humains.

L'empathie artificielle a pour but d'améliorer la coopération avec les agents conversationnels. Elle se sert de la détection des émotions pour simuler certains aspects de l'empathie humaine, c'est-à-dire permettre à l'agent conversationnel de se comporter comme s'il se mettait à la place de l'utilisateur. Cela a pour conséquence un anthropomorphisme affectif, qui peut se révéler utile en vue de la finalité recherchée, mais aussi potentiellement néfaste. Les conséquences potentielles pour les êtres humains incluent diverses réactions d'attachement, de culpabilité ou de confiance, que ce soit envers l'agent conversationnel ou d'autres êtres humains. Définir des limites aux simulations émotionnelles des chatbots est une tâche qui dépend des contextes d'application comme des vulnérabilités humaines. Par exemple, un chatbot « triste » pourrait amplifier une dépression, mais aussi soulager une douleur.

PRINCIPE DE CONCEPTION 8

RESPECTER LA PROPORTIONNALITÉ LORS DU DÉPLOIEMENT DES TECHNOLOGIES D'INFORMATIQUE AFFECTIVE DANS LES CHATBOTS

Pour réduire la projection spontanée d'émotions sur l'agent conversationnel et l'attribution d'une intériorité à ce système, le fabricant devrait respecter la proportionnalité et l'adéquation entre les finalités recherchées et le déploiement des technologies d'informatique affective, notamment la détection du comportement émotionnel des personnes et l'empathie artificielle du chatbot. Il doit également informer l'utilisateur des biais éventuels de l'anthropomorphisme.

PRINCIPE DE CONCEPTION 9

ADAPTER LES AGENTS CONVERSATIONNELS AUX CODES CULTURELS

Les concepteurs des chatbots devraient adapter les agents conversationnels aux codes culturels d'expression des émotions dans différentes parties du monde.

PRINCIPE DE CONCEPTION 10

COMMUNIQUER SUR LES CAPACITÉS DES AGENTS CONVERSATIONNELS AFFECTIFS

Dans leur communication sur les agents conversationnels émotionnels, les concepteurs doivent veiller à expliquer les limites et les capacités réelles de ces systèmes aux utilisateurs pour ne pas donner prise à des surinterprétations de ces simulations affectives.

5. ÉVALUATION DES AGENTS CONVERSATIONNELS

Un agent conversationnel fournit une réponse en appliquant des stratégies de dialogue qui dépendent de l'apprentissage effectué par des modèles de manière automatisée. Les modèles les plus avancés utilisent de grands corpus de données. L'évaluation de tels systèmes, par essence dynamique, fait partie de la démarche « éthique par conception » (section III.1). Elle est difficile au moins sur deux plans : a) la prédictibilité du comportement langagier de l'utilisateur ; b) la prédictibilité des stratégies de dialogue, ce qui contribue à la difficulté de reproduire le comportement du système.

Cette incertitude théorique et pratique va de pair avec les techniques d'apprentissage qui procurent aux systèmes leur grande efficacité. L'évaluation du comportement des systèmes apprenants est donc un enjeu de première importance.

L'enjeu d'évaluation est particulièrement présent pour les systèmes d'apprentissage adaptatif, même si ceux-ci sont aujourd'hui encore assez rares parmi les chatbots commercialisés. Dans le cas des agents conversationnels, la qualité d'un dialogue est souvent mesurée par l'engagement de l'utilisateur, c'est-à-dire sa persistance à poursuivre le dialogue avec le chatbot. À travers l'apprentissage par renforcement, des métriques qui servent à trouver les phrases permettant d'optimiser l'engagement de l'utilisateur font intervenir la durée des échanges, mais aussi des marqueurs paralinguistiques de satisfaction ou d'intérêt de l'utilisateur (rire, sourire, hésitation, hochement de tête, etc.).

APPRENTISSAGE PAR RENFORCEMENT DANS LES CHATBOTS

En apprentissage par renforcement, on dit que l'agent conversationnel interagit avec « l'environnement » pour trouver la solution optimale. Ce type d'apprentissage se distingue par son côté interactif et itératif : au cours de l'interaction, plusieurs solutions sont essayées et, en fonction des réactions de l'environnement dans lequel évolue l'agent conversationnel, des adaptations sont effectuées afin d'aboutir à la meilleure stratégie. Le but est d'apprendre, à partir d'expériences, ce qu'il convient de faire en différentes situations, de façon à optimiser une récompense globale quantitative au cours du temps. Certains chercheurs tentent de prouver que l'apprentissage par renforcement serait suffisant pour rendre compte de toutes les manifestations de l'intelligence humaine[59] ; ce débat d'envergure mérite d'être approfondi dans le contexte des agents conversationnels.

[59] D. Silver, S. Singh, D. Precup, and R. S. Sutton, "Reward is enough". Artificial Intelligence 299 (2021) 103535.

Ces stratégies de dialogue peuvent s'avérer éthiquement inacceptables. Par exemple, s'il est observé que, statistiquement, les utilisateurs tendent à répondre aux injures qui leur sont adressés, un agent conversationnel pourrait injurier son utilisateur afin de maximiser son engagement dans le dialogue.

UN CHATBOT ADAPTATIF : TAY

En avril 2016, le chatbot Tay de Microsoft, doté de la capacité d'apprendre de façon adaptative à partir de ses interactions avec les internautes, a appris lors de sa mise en ligne à tenir des propos racistes. DeepCom, un autre chatbot développé par Microsoft China en 2019 afin de commenter des nouvelles sur les réseaux sociaux, a été reconnu par les chercheurs eux-mêmes « être susceptible de générer des contenus biaisés, voire de la propagande, suite à de fortes réactions dans la communauté de recherche »[60].

Sur le plan de la sécurité, le développement des « deepfakes » à l'aide des systèmes apprenants peut poser un problème pour les entreprises qui se fient aux procédures orales ou informelles pour leur fonctionnement. Des chatbots utilisant des « deepfakes » peuvent être utilisés afin de commander de fausses transactions financières.

MESURES JURIDIQUES RELATIVES AUX « DEEPFAKES »

L'article 52 de la proposition de règlement européen sur l'intelligence artificielle prévoit que le fabricant d'un système d'intelligence artificielle « qui génère ou manipule du contenu image, audio ou vidéo, qui ressemble sensiblement à des personnes, des objets ou des lieux ou d'autres entités ou événements existants et qui apparaîtraient à tort à une personne comme étant authentiques ou véridiques (« deep fake »), doit divulguer que le contenu a été généré ou manipulé artificiellement ». Le manquement à une telle obligation serait sanctionné par une amende (article 71).

Des erreurs sont inévitables lorsqu'un système apprenant classifie une donnée qui ne ressemble pas à celles contenues dans le corpus utilisé pendant son apprentissage. Dans le cas des agents conversationnels, cela recouvre par exemple les homophones, homographes, homonymes ou d'autres exemples d'ambiguïté linguistique. Des pirates informatiques peuvent instrumentaliser cette propriété d'instabilité inhérente aux systèmes apprenants, par exemple pour diriger les chatbots vers des choix de parole inopportuns, des conseils nuisibles ou des dysfonctionnements.

Les chatbots vont utiliser des corpus de données de taille toujours plus grande. Le déploiement des réseaux de neurones de type transformeur étant récent, on ne dispose pas de données expérimentales permettant d'évaluer leurs effets. Par exemple, GPT-3 n'est pas capable d'exclure systématiquement de ses résultats les propos racistes, sexistes et haineux. C'est un problème technique complexe. Il est donc nécessaire de développer des méthodes d'évaluation adaptées à ces réseaux de neurones de très grande taille.

QUESTION DE RECHERCHE 10

DÉVELOPPER DES MÉTHODES D'ÉVALUATION ADAPTÉES AUX AGENTS CONVERSATIONNELS

Les autorités publiques et les acteurs privés doivent soutenir des recherches sur l'évaluation des agents conversationnels pendant leur utilisation et proposer de nouveaux tests adaptés au contexte d'utilisation.

QUESTION DE RECHERCHE 11

ÉTUDIER LES CAPACITÉS DES RÉSEAUX DE NEURONES DE TYPE TRANSFORMEUR POUR LE DIALOGUE

Au vu de leurs capacités de traitement et de génération de langage, il est nécessaire de soutenir des recherches sur les agents conversationnels utilisant les réseaux de neurones de type transformeur, notamment pour l'évaluation de leur conformité aux valeurs éthiques.

[60] arXiv:1909.11974.

IV. LISTE DES PRÉCONISATIONS, PRINCIPES DE CONCEPTION ET QUESTIONS DE RECHERCHE

PRÉCONISATIONS

PRÉCONISATION 1 : RÉDUIRE LA PROJECTION DE QUALITÉS MORALES SUR UN AGENT CONVERSATIONNEL

Pour réduire la projection spontanée de qualités morales sur l'agent conversationnel et l'attribution d'une responsabilité à ce système, le fabricant doit limiter sa personnification et informer l'utilisateur des biais éventuels issus de l'anthropomorphisation de l'agent conversationnel.

PRÉCONISATION 2 : AFFIRMER LE STATUT DES AGENTS CONVERSATIONNELS

Toute personne qui communique avec un agent conversationnel doit être informée de manière adaptée, claire et compréhensible du fait qu'elle dialogue avec une machine.

PRÉCONISATION 3 : PARAMÉTRER L'IDENTITÉ DES AGENTS CONVERSATIONNELS

Pour éviter les biais, notamment de genre, le choix par défaut des caractéristiques d'un agent conversationnel à usage public (nom, pronoms personnels, voix) doit être effectué de façon équitable à chaque fois que cela est possible. S'agissant des agents conversationnels personnels à usage privé ou domestique, l'utilisateur doit pouvoir modifier ces choix par défaut.

PRÉCONISATION 4 : TRAITER LES INSULTES

S'il est impossible d'exclure les situations où l'utilisateur profère des insultes envers un agent conversationnel, le fabricant doit les prévoir et définir des stratégies de réponse spécifiques. Notamment, l'agent conversationnel ne devrait pas répondre aux insultes par des insultes et ne pas les rapporter à une autorité. Le fabricant d'un agent conversationnel apprenant doit veiller à exclure de telles phrases du corpus d'apprentissage.

PRÉCONISATION 5 : INFORMER SUR LA MANIPULATION À DESSEIN

Dans l'hypothèse où l'agent conversationnel a été programmé de manière à pouvoir influencer le comportement de l'utilisateur dans le cadre de sa finalité, le fabricant doit informer l'utilisateur de l'existence de cette capacité et recueillir son consentement qu'il doit pouvoir à tout moment retirer. Le fabricant d'un agent conversationnel influenceur doit permettre aux utilisateurs d'être informés sur la nature, l'origine et les modalités de diffusion des contenus, et leur demander d'être vigilants avant de repartager ces contenus.

PRÉCONISATION 6 : ÉVITER LA MANIPULATION MALVEILLANTE

Le fabricant doit veiller à éliminer les manipulations malveillantes ou les menaces de la part de l'agent conversationnel. L'utilisateur doit avoir la capacité de signaler certaines expressions non souhaitées en vue d'une modification de l'agent conversationnel par le concepteur.

PRÉCONISATION 7 : ENCADRER L'USAGE DES CHATBOTS DANS LES JOUETS

Dans le domaine ludique, tout particulièrement pour la petite enfance, les autorités publiques se doivent d'évaluer les conséquences des interactions avec des chatbots, susceptibles de modifier le comportement des enfants. Les autorités publiques doivent encadrer l'utilisation des agents conversationnels auprès d'enfants au regard de l'impact de cette interaction sur le développement langagier, émotionnel et culturel de l'enfant.

PRÉCONISATION 8 : RESPECTER LES PERSONNES VULNÉRABLES

Dans le cas du dialogue entre un agent conversationnel et une personne vulnérable, le fabricant de l'agent conversationnel doit veiller à respecter la dignité et l'autonomie de cette personne. Notamment dans le domaine médical, il est nécessaire, dès l'étape de conception des agents conversationnels, d'éviter la confiance excessive en ces systèmes de la part du patient et de veiller à lever la confusion entre l'agent conversationnel et le médecin qualifié.

PRÉCONISATION 9 : ANALYSER LES EFFETS DES AGENTS CONVERSATIONNELS COUPLÉS À DES MESURES PHYSIOLOGIQUES

Dans les cas de couplage des agents conversationnels avec des mesures physiologiques (« Quantified Self »), les concepteurs doivent mener des analyses portant sur les risques de dépendance. Les autorités publiques doivent encadrer l'utilisation de ces systèmes au regard de leur impact sur l'autonomie de la personne.

PRÉCONISATION 10 : DÉFINIR LES RESPONSABILITÉS POUR L'USAGE DES AGENTS CONVERSATIONNELS DANS LE MILIEU PROFESSIONNEL

Le fabricant doit prévoir des mécanismes de contrôle et d'audit afin de faciliter l'attribution de responsabilités au regard du bon fonctionnement ou du dysfonctionnement de l'agent conversationnel dans le milieu professionnel, notamment étudier leurs effets secondaires ou non-intentionnels.

PRÉCONISATION 11 : MENER UNE RÉFLEXION SOCIÉTALE AVANT TOUTE RÉGLEMENTATION DES « DEADBOTS »

À la suite d'une réflexion éthique approfondie à l'échelle de toute la société, le législateur doit adopter une réglementation spécifique concernant les agents conversationnels qui imitent la parole des personnes décédées.

PRÉCONISATION 12 : ENCADRER TECHNIQUEMENT LES « DEADBOTS »

Les concepteurs de « deadbots » doivent respecter la dignité de la personne humaine qui ne s'éteint pas avec la mort, tout en veillant à préserver la santé mentale des utilisateurs de tels agents conversationnels. Des règles doivent être définies et respectées concernant notamment le consentement de la personne décédée, le recueil et à la réutilisation de ses données, le temps de fonctionnement d'un tel chatbot, le lexique utilisé, le nom qui lui est attribué ou encore les conditions particulières de son utilisation.

PRÉCONISATION 13 : ENCADRER LE DÉPLOIEMENT DES CHATBOTS « ANGE GARDIEN »

Les autorités publiques doivent encadrer l'utilisation des agents conversationnels de type « ange gardien », qui sont capables de protéger les données d'une personne, afin de limiter le paternalisme et respecter l'autonomie humaine.

PRINCIPES DE CONCEPTION DES CHATBOTS

PRINCIPE DE CONCEPTION 1 : « ÉTHIQUE PAR CONCEPTION » DES AGENTS CONVERSATIONNELS

Les concepteurs d'un agent conversationnel doivent analyser, en phase de conception, chacun des choix technologiques susceptibles de provoquer des tensions éthiques. Si une tension potentielle est identifiée, ils doivent envisager une solution technique visant à diminuer ou à faire disparaître la tension éthique, puis évaluer cette solution dans des contextes d'usage réalistes.

PRINCIPE DE CONCEPTION 2 : RÉDUIRE LES BIAIS DANS LE LANGAGE

Pour réduire les biais dans le langage et tenter d'éviter les effets de discrimination, notamment culturels, les concepteurs doivent mettre en œuvre une solution technique à trois niveaux : dans la mise en œuvre de l'algorithme, dans la sélection des paramètres d'optimisation et dans le choix des données d'apprentissage et de validation pour les différents modules des agents conversationnels.

PRINCIPE DE CONCEPTION 3 : ÉNONCER LES FINALITÉS DU CHATBOT

Le concepteur doit veiller à ce qu'un agent conversationnel communique sa finalité à l'utilisateur de manière claire et aisément compréhensible au moment adéquat, par exemple au début ou à la fin de chaque conversation.

PRINCIPE DE CONCEPTION 4 : TRANSPARENCE ET TRAÇABILITÉ DU CHATBOT

L'agent conversationnel devrait pouvoir sauvegarder le contenu du dialogue à des fins de preuve en cas de litige. Cela introduit une tension entre les données privées et celles sauvegardées pour assurer la transparence des décisions du chatbot. L'architecture des chatbots, les connaissances utilisées et les stratégies de dialogue doivent être accessibles à toutes les parties concernées, notamment afin de faciliter le traitement d'éventuels problèmes juridiques. Cette préconisation pourrait déboucher sur une mesure d'ordre réglementaire définissant les conditions précises de son application.

PRINCIPE DE CONCEPTION 5 : TRAITEMENT DES DONNÉES COLLECTÉES PAR LES AGENTS CONVERSATIONNELS

À l'image de l'encadrement existant des données de santé, il est nécessaire d'élaborer des règles éthiques et juridiques pour la collecte, le stockage et l'utilisation des traces linguistiques des interactions entre les agents conversationnels dans le respect du RGPD.

PRINCIPE DE CONCEPTION 6 : INFORMER SUR LES CAPACITÉS DES AGENTS CONVERSATIONNELS

Dans un souci de transparence, l'utilisateur doit être informé de manière adaptée, claire et compréhensible, oralement ou par écrit, de la collecte des données, des capacités de l'agent conversationnel liées à l'adaptation aux données qu'il collecte en cours d'utilisation et au profilage.

PRINCIPE DE CONCEPTION 7 : FAVORISER L'EXPLICABILITÉ DU COMPORTEMENT DU CHATBOT

Les concepteurs doivent développer des solutions favorisant la compréhension par les utilisateurs du comportement des agents conversationnels.

PRINCIPE DE CONCEPTION 8 : RESPECTER LA PROPORTIONNALITÉ LORS DU DÉPLOIEMENT DES TECHNOLOGIES D'INFORMATIQUE AFFECTIVE DANS LES CHATBOTS

Pour réduire la projection spontanée d'émotions sur l'agent conversationnel et l'attribution d'une intériorité à ce système, le fabricant devrait respecter la proportionnalité et l'adéquation entre les finalités recherchées et le déploiement des technologies d'informatique affective, notamment la détection du comportement émotionnel des personnes et l'empathie artificielle du chatbot. Il doit également informer l'utilisateur des biais éventuels de l'anthropomorphisme.

PRINCIPE DE CONCEPTION 9 : ADAPTER LES AGENTS CONVERSATIONNELS AUX CODES CULTURELS

Les concepteurs des chatbots devraient adapter les agents conversationnels aux codes culturels d'expression des émotions dans différentes parties du monde.

PRINCIPE DE CONCEPTION 10 : COMMUNIQUER SUR LES CAPACITÉS DES AGENTS CONVERSATIONNELS AFFECTIFS

Dans leur communication sur les agents conversationnels émotionnels, les concepteurs doivent veiller à expliquer les limites et les capacités réelles de ces systèmes aux utilisateurs pour ne pas donner prise à des surinterprétations de ces simulations affectives.

QUESTIONS DE RECHERCHE

1.QUESTION DE RECHERCHE : RECONNAÎTRE AUTOMATIQUEMENT LES PROPOS NON DÉSIRABLES

Il est nécessaire de développer des méthodes de caractérisation automatique par les agents conversationnels de propos non désirables, notamment des insultes.

2.QUESTION DE RECHERCHE : ETUDIER LES MENSONGES PROFÉRÉS PAR UN AGENT CONVERSATIONNEL

Il est nécessaire d'étudier des méthodes pour soustraire l'agent conversationnel à la projection de qualités morales à travers une mise en récit de ses actions explicitement différente de celle qui caractérise les mensonges proférés par les êtres humains.

QUESTION DE RECHERCHE 3 : ÉVALUER LES EFFETS ÉDUCATIFS INÉDITS DES CHATBOTS

Dans le domaine de l'éducation, tout particulièrement pour les enfants vulnérables et pour la petite enfance, les autorités publiques se doivent d'évaluer les conséquences des interactions entre les élèves et le chatbot.

QUESTION DE RECHERCHE 4 : ÉTUDIER LES EFFETS DES AGENTS CONVERSATIONNELS SUR L'ORGANISATION DU TRAVAIL

Les autorités publiques et les acteurs privés doivent soutenir des recherches empiriques sur les effets organisationnels de l'introduction d'agents conversationnels dans les équipes selon les secteurs professionnels.

QUESTION DE RECHERCHE 5 : ÉTUDIER LES EFFETS DE L'UTILISATION DES CHATBOTS À LONG TERME

Les autorités publiques et les acteurs privés doivent investir dans des recherches sur les effets à long terme et les conséquences de l'utilisation des agents conversationnels sur l'être humain et la société. Tous les acteurs de la société doivent rester vigilants quant aux effets futurs des agents conversationnels sur les croyances, opinions et décisions des utilisateurs, notamment des effets de masse, et éviter de considérer cette technologie comme neutre ou dépourvue de signification éthique et politique.

QUESTION DE RECHERCHE 6 : ÉTUDIER L'IMPACT ENVIRONNEMENTAL DES AGENTS CONVERSATIONNELS

Les autorités publiques et les acteurs privés doivent mener des études sur l'impact énergétique et environnemental de la technologie des agents conversationnels.

QUESTION DE RECHERCHE 7 : DÉVELOPPER LES MÉTHODOLOGIES « ÉTHIQUE PAR CONCEPTION » POUR LES CHATBOTS

Les autorités publiques doivent soutenir des recherches afin d'élaborer des méthodologies « éthique par conception » adaptées au développement des agents conversationnels.

QUESTION DE RECHERCHE 8 : REPRODUCTIBILITÉ DES PROPOS DES AGENTS CONVERSATIONNELS

La reproductibilité nécessite de mémoriser les données, mais aussi de définir la bonne mesure de répétition dans les propos du chatbot. Ces questions doivent être étudiées.

QUESTION DE RECHERCHE 9 : ÉTUDIER LES EFFETS DES CHATBOTS SUR LE COMPORTEMENT ÉMOTIONNEL DES ÊTRES HUMAINS

Dans le domaine émergeant des agents conversationnels empathiques, les concepteurs doivent développer des recherches et réaliser des analyses de risque relatives aux répercussions éventuelles de ces systèmes sur le comportement émotionnel de l'utilisateur humain, notamment sur le long terme.

QUESTION DE RECHERCHE 10 : DÉVELOPPER DES MÉTHODES D'ÉVALUATION ADAPTÉES AUX AGENTS CONVERSATIONNELS

Les autorités publiques et les acteurs privés doivent soutenir des recherches sur l'évaluation des agents conversationnels pendant leur utilisation et proposer de nouveaux tests adaptés au contexte d'utilisation.

QUESTION DE RECHERCHE 11 : ETUDIER LES CAPACITÉS DES RÉSEAUX DE NEURONES DE TYPE TRANSFORMEUR POUR LE DIALOGUE

Au vu de leurs capacités de traitement et de génération de langage, il est nécessaire de soutenir des recherches sur les agents conversationnels utilisant les réseaux de neurones de type transformeur, notamment pour l'évaluation de leur conformité aux valeurs éthiques.

ANNEXE 1 : CONSENTEMENT

Si l'on place le débat sur les agents conversationnels dans une perspective informatique et libertés, le principal enjeu concerne l'information communiquée à l'utilisateur pour indiquer clairement la manière dont ses données personnelles sont susceptibles d'être traitées pour obtenir son consentement, et le cas échéant, offrir des possibilités de limiter ces traitements ou de s'y opposer.

LE CONSENTEMENT DANS LE RGPD ET LA LOI DU 6 JANVIER 1978

La question de la protection des données personnelles a pris une acuité particulière avec le développement du numérique, l'explosion du traitement des données et la mise à la disposition de services gratuits en contrepartie de l'utilisation des données. La protection de la collecte des données personnelles, qui constitue un aspect de la protection de la vie privée, a été prise en compte très tôt par le droit français avec la loi du 6 janvier 1978[61], dite loi informatique et libertés. Elle a fait l'objet d'un premier encadrement par le droit de l'Union européenne en 2002 en ce qui concerne les communications[62]. Elle est désormais encadrée par le règlement européen du 27 avril 2016[63], dit RGPD, qui, notamment à travers ses chapitres II et III[64], constitue un ensemble très protecteur des droits de la personne.

Le consentement est l'une des six bases légales qui autorisent la mise en œuvre de traitements de données personnelles, les autres étant l'obligation légale, le contrat (relations contractuelles ou précontractuelles), la mission d'intérêt public, la sauvegarde des intérêts vitaux, l'intérêt légitime (par exemple, des opérations de prospection commerciale auprès de clients d'une société sans conclusion de contrat).

Conformément à l'article 12 du RGPD, l'utilisateur doit être informé de manière concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples, de la politique de confidentialité. Cette information doit porter sur la finalité de la collecte, l'obligation ou non dans laquelle la personne se trouve de fournir les données et les conséquences éventuelles d'un refus, sur les données collectées, le responsable de traitement, les coordonnées du délégué à la protection des données, les destinataires des données, les sous-traitants éventuels, le transfert éventuel hors UE (États tiers, garanties appropriées prévues pour encadrer le transfert), la durée de conservation, l'exercice des droits des personnes concernées sur les données détenues à leur sujet avec possibilité d'un droit d'accès aux données, de rectification ou d'effacement de celles-ci...

L'utilisateur d'un agent conversationnel, en tant que consommateur, doit donner un consentement libre, spécifique, éclairé et univoque (sauf exception prévue par la loi). Il peut retirer son consentement. L'article 7 du RGPD dispose que le responsable du traitement doit s'assurer qu'il est aussi simple pour la personne concernée de retirer que de donner son consentement, et qu'elle peut le retirer à tout moment. L'existence d'un retrait facile, dont doit être informé l'utilisateur, est une garantie pour un consentement valable. Précisons que le RGPD interdit le traitement de données biométriques (par exemple les gabarits ou modèles de voix) qui sont considérées comme des données sensibles, sauf certaines exceptions limitativement énumérées.

Le contrôle de la protection des données personnelles relève d'un régulateur national, la CNIL, qui veille au respect du RGPD et de la loi informatique et libertés[65], en formulant notamment des avis et des mises en demeure et en appliquant des sanctions[66] sous le contrôle du Conseil d'État. Alors que le juge national et la Cour de justice de l'Union européenne développent progressivement la jurisprudence sur la protection des données, se posent de nombreuses questions sur la qualité du consentement, sur son sens et sur les conditions dans lesquelles il est recueilli (lisibilité, clarté et précision des clauses). De nombreuses tensions existent entre le droit et la collecte effective des données, incitant à une réflexion permanente dans ce domaine.

Comment peut-on consentir à des clauses qui ne sont pas facilement explicables et peuvent être nombreuses ? Comment recueillir effectivement le consentement parental en ce qui concerne les mineurs[67]? La demande constante du consentement ne peut-elle conduire à une fiction juridique ? Faudrait-il revoir les modalités d'information, en innovant s'agissant de la présentation des clauses engageant le consentement numérique ? En recourant au jeu, au design, aux images ? Comment faire comprendre les conséquences du consentement face à une proposition attractive ?

---

[61] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et diverses dispositions concernant la protection des données à caractère personnel.

[62] Directive 2002/58/CE du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques (directive e-privacy) qui devrait être remplacée par le règlement e-privacy.

[63] 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, et abrogeant la directive 95/46/CE (règlement général sur la protection des données).

[64] Principes et droits de la personne concernée.

[65] La loi informatique et libertés dans sa version actuelle contient un certain nombre de dispositions reprises du RGPD, comporte des précisions par rapport à celui-ci, incorpore des articles de la loi du 7 octobre 2016 pour une République numérique et transpose la directive e-privacy du 12 juillet 2002.

[66] Voir sanction de la CNIL du 21 janvier 2019 à l'encontre de Google : un des motifs est que les utilisateurs ne sont pas toujours en mesure d'appréhender les incidences des traitements en cause sur leur vie privée ; l'exigence d'un consentement éclairé et d'un consentement spécifique et univoque n'est pas respectée. Le recours contre cette décision a été rejeté par une décision du Conseil d'État du 9 juin 2020 Google LLC n° 430810.

[67] Moins de seize ans (article 8 du RGPD) ; moins de quinze ans article (article 7-1 de la loi informatique et libertés).

Quel est le sens du consentement lorsqu'on recourt à un agent conversationnel ? La notion de transparence de l'information peut paraître illusoire : elle est en effet fonction des données collectées, des fonctions de cet agent ou de son mode d'expression (écrit ou oral). L'information ne peut être standardisée mais adaptée. Délivrer une information satisfaisante lorsqu'il s'agit d'agents conversationnels se substituant à des personnes physiques pour des services après-vente ou fournissant des services promotionnels est particulièrement compliqué à cause des tensions entre les exigences de transparence et d'efficacité. En effet, le temps est limité et le consommateur, pressé, n'est pas spécialement intéressé par la protection de ses données dès lors qu'il entend avant tout résoudre un problème technique ou trouver une proposition commerciale. L'entreprise est tentée de viser la rapidité plutôt que l'exhaustivité pour être attractive et efficace, avec le risque de perdre la confiance de l'utilisateur en ne lui expliquant pas la finalité de l'agent conversationnel.

La question du consentement peut être particulièrement sensible s'il s'agit d'un agent conversationnel installé à domicile ou embarqué sur un périphérique mobile et destiné à être un confident ou un assistant, avec lequel s'engagent des conversations très intimes, avec de très importants risques d'atteinte à la vie privée. Si l'information doit être proportionnée, elle doit cependant être délivrée, dans cette hypothèse, de manière à ce que le consentement soit éclairé. Or, il n'est pas facile d'évaluer si l'utilisateur a compris les enjeux de l'information qui lui est délivrée.

ANNEXE 2 : APPEL À CONTRIBUTION

Pour répondre à la saisine du Premier ministre, le CNPEN a ouvert en juillet 2020 un appel à contribution sur les enjeux éthiques des agents conversationnels avec un triple objectif : enrichir sa propre réflexion grâce aux réponses reçues, sensibiliser les différents acteurs du numérique et l'ensemble des citoyens aux questions éthiques que soulèvent ces technologies. La consultation a été diffusée, en français et en anglais, vers les milieux académiques et universitaires, institutionnels ou industriels, en France et à l'étranger.

Quatre-vingt-seize contributions, individuelles ou collectives, rendues à titre personnel ou professionnel, ont été recueillies entre juillet 2020 et janvier 2021. Elles émanent pour partie de personnes issues des secteurs universitaires ou académiques : étudiants, responsables pédagogiques, enseignants ou chercheurs, français ou étrangers, dont une part importante des sciences juridiques et informatiques. Plusieurs contributions proviennent par ailleurs de personnes travaillant dans les secteurs industriels (télécommunications, conception de chatbots…) ou tertiaires (banque, data analysis, conseil). Quelques contributions viennent enfin des secteurs de la santé, de la justice ou de la régulation du numérique.

Le questionnaire a également été utilisé comme outil pédagogique par des enseignants et comme base de discussion au sein d'entreprises pour nourrir la réflexion autour de l'éthique des agents conversationnels. Ces utilisations, qui n'avaient pas nécessairement été anticipées par le comité, ont contribué à conforter l'idée que cette démarche de consultation était pertinente et utile ; elles témoignent en effet d'un besoin d'outils concrets de formation, d'acculturation et de sensibilisation du public.

Les questions posées mettaient volontairement en exergue les tensions éthiques liées aux différents usages des chatbots, illustrées par des exemples concrets. Les contributions reflètent l'hétérogénéité des connaissances de ces technologies ainsi qu'un niveau de perception inégal des enjeux éthiques.

Par exemple, pour certains contributeurs, l'anthropomorphisation des agents conversationnels est nécessaire à leur engagement dans l'interaction. Pour d'autres, cette tendance doit être inhibée autant que possible : l'emploi même de la première personne par le chatbot devrait alors être proscrit. Par ailleurs, les questions relatives aux erreurs, au mensonge, aux émotions, ont suscité des réponses contrastées, impliquant des choix moraux enracinés dans des croyances, imaginaires et traditions de différentes origines.

Bien que les contributions rassemblées ne puissent être considérées comme représentatives, certaines réponses sont manifestement consensuelles. Par exemple, les questions relatives au libre choix soulignent l'importance du principe de transparence et du respect de l'autonomie de l'utilisateur. Aux yeux des répondants, il revient donc aux concepteurs de trouver les moyens techniques et la conception adaptée pour permettre la mise en œuvre pratique de ce principe.

Il était intéressant de distinguer les réponses formulées du point de vue des concepteurs de chatbots et celles se situant davantage du côté des utilisateurs. Les réponses à propos de la confiance dans les chatbots sont en ce sens particulièrement intéressantes. Les concepteurs considèrent que les situations où le chatbot ne peut répondre correctement à l'utilisateur ne peuvent être évitées. En revanche, les utilisateurs ne tolèrent pas les erreurs, qu'ils assimilent à des manquements dans la conception du chatbot. Sans connaissance technique et sans expérience préalable, il est difficile de saisir correctement les causes du comportement du chatbot.

LES ENJEUX ÉTHIQUES DES AGENTS CONVERSATIONNELS

APPEL À CONTRIBUTIONS

ENVOI DES RÉPONSES À L'ADRESSE CNPEN-CONSULTATION-CHATBOTS@CCNE.FR

Le Comité national pilote d'éthique du numérique (CNPEN) a été créé en décembre 2019 à la demande du Premier ministre. Constitué de 27 membres, ce comité réunit des spécialistes du numérique, des philosophes, des médecins, des juristes et des membres de la société civile. L'une des trois saisines soumises par le Premier ministre au CNPEN concerne les enjeux éthiques des agents conversationnels, appelés communément chatbots, qui communiquent avec l'utilisateur humain par la voix ou par écrit. Ce travail du CNPEN vient en prolongation des travaux initiés par la CERNA, Commission d'éthique de la recherche en sciences et technologies du numérique de l'alliance Allistene.

Dans ce document, nous sollicitons l'avis des lecteurs en posant des questions sur les enjeux éthiques liés aux chatbots. Chacun est invité à répondre soit à quelques questions de son choix soit à l'ensemble des questions posées.

Répondez-vous à ce questionnaire :

- À titre personnel (préciser vos nom et prénom si vous le souhaitez)

- Au titre de vos activités professionnelles ou au nom d'une organisation :`

- *Chercheur ou Institut de recherche (préciser le nom de votre institution)*
- *Société ou groupe de sociétés (préciser laquelle)*
- *Association de consommateurs ou assimilé (préciser laquelle)*
- *Autorité publique (préciser laquelle)*
- *Consultant professionnel*
- *Think thank (préciser lequel)*
- *Autre :*

Objectifs de ce document :

Le Comité national pilote d'éthique du numérique (CNPEN), créé en décembre 2019 sous l'égide du CCNE pour les sciences de la vie et de la santé, a été saisi par le Premier ministre pour élaborer en particulier un avis sur les enjeux éthiques des agents conversationnels (chatbots). Il a aussi dans ses objectifs « d'engager une discussion collective pour développer une approche partagée des innovations présentes et futures. Cette dimension est fondamentale pour s'assurer que la technique et l'innovation continuent à servir le bien commun. ». C'est pourquoi le Comité engage une consultation des parties prenantes et des citoyens avec pour objectifs de les sensibiliser aux enjeux éthiques et d'enrichir sa réflexion.

Utilisation et protection de vos données personelles :

Les données personnelles demandées (adresse mail, nom, prénom, profession, institution de rattachement) ou celles que vous pourriez fournir spontanément dans votre réponse au questionnaire ne seront traitées que si elles sont utiles à l'analyse et à la réflexion du comité. Toutes les données personnelles récoltées seront conservées sur les serveurs du CCNE ou de ses prestataires. Elles seront traitées de manière confidentielle uniquement par le personnel du CNPEN ou les membres du groupe de travail du CNPEN sur les agents conversationnels ; elles ne seront pas traitées de manière automatisées. Elles seront conservées au maximum dix-huit mois après la clôture de la consultation et jusqu'à douze mois après la publication de l'avis du comité.

Les résultats de cette analyse nourriront l'avis du comité sur les agents conversationnels, qui sera rendu public. Les contributions n'y seront pas citées nommément sans l'accord explicite de leurs auteurs.

Dans les conditions définies par la Loi Informatique et Libertés du 6 janvier 1978 et par le Règlement Européen sur la Protection des Données Personnelles entré en vigueur le 25 mai 2018, chaque contributeur bénéficie d'un droit d'accès aux données le concernant, de rectification, d'interrogation, de limitation, de portabilité et d'effacement. Chaque contributeur peut également, pour des motifs légitimes, s'opposer au traitement de ses données personnelles. Le contributeur peut exercer l'ensemble des droits mentionnés ci-dessus en s'adressant au CNPEN à l'adresse : cnpen-consultation-chatbots@ccne.fr.

INTRODUCTION

QU'EST-CE QU'UN AGENT CONVERSATIONNEL ?

Un agent conversationnel, appelé communément chatbot, est un programme informatique qui interagit avec son utilisateur dans la langue naturelle de celui-ci. Ces termes regroupent tant les agents vocaux que les chatbots écrits.

Le plus souvent, un agent conversationnel ne constitue pas une entité indépendante mais il est intégré au sein d'un système ou d'une plate-forme numérique, comme un smartphone ou une enceinte vocale. Sur le plan de l'apparence visuelle, les chatbots peuvent aussi être intégrés à un agent conversationnel animé, représenté en deux ou trois dimensions sur un écran, voire faire partie d'un robot social, y compris humanoïde. Le dialogue avec l'utilisateur ne représente alors qu'une des fonctions du système global.

L'histoire des agents conversationnels prend ses origines dans le jeu de l'imitation d'Alan Turing. La compréhension du langage intéresse Turing dans la mesure où elle se manifeste à travers des réponses qui paraissent intelligibles et sensées à un examinateur (« test de Turing »). Dès 1991, un concours annuel est organisé afin de soutenir le développement de chatbots capables de passer le test de Turing.

Le premier agent conversationnel de l'histoire de l'informatique est le programme ELIZA de Joseph Weizenbaum[70], qui est aussi l'un des premiers leurres conversationnels. ELIZA simule un dialogue écrit avec un psychothérapeute rogérien en reformulant tout simplement la plupart des répliques de l'utilisateur « patient » sous forme de questions. Aujourd'hui, l'expression « effet ELIZA » désigne la tendance à assimiler de manière inconsciente le dialogue avec un ordinateur à celui avec un être humain.

D'UN POINT DE VUE TECHNIQUE, COMMENT ÇA MARCHE ?

La conception et le fonctionnement d'un agent conversationnel se divisent en plusieurs modules de traitement automatique du langage naturel (TALN) : schématiquement, un chatbot peut inclure des modules de reconnaissance de la parole (pour les agents conversationnels vocaux), de traitement sémantique (hors et en contexte), de gestion de l'historique du dialogue, de gestion des stratégies de dialogue, d'accès aux ontologies, de gestion des accès aux connaissances externes (base de données ou internet), de génération de langage et de synthèse de la parole (pour les agents conversationnels vocaux).

Un agent conversationnel suit des règles décidées et transposées en code par des concepteurs humains ou obtenues par apprentissage. Les chatbots apprenants, par exemple Xiaoice de Microsoft Chine[71], sont aujourd'hui encore assez rares parmi les applications commercialisées, mais leur proportion n'aura de cesse de croître avec l'avancement de la maîtrise de cette technologie.

Ces dernières années, développer soi-même un chatbot rudimentaire ou dédié à une seule tâche est devenu relativement facile grâce à la disponibilité de nombreux outils de conception, comme "LiveEngage", "Chatbot builder", "Passage.ai", "Plato Research Dialogue System", etc.

QUELQUES DÉFIS DE RECHERCHE CONCERNANT LA CONCEPTION DES AGENTS CONVERSATIONNELS

- Apprendre de manière adaptative en faisant évoluer la base de connaissances en cours d'utilisation.
- Être capable de converser librement sur des sujets génériques.
- Saisir le « sens commun », le caractère ironique ou le sens au « second degré » d'un énoncé
- Mettre en place une stratégie de dialogue.

DÉTECTER LES ÉMOTIONS ET LES INTENTIONS DE L'UTILISATEUR. QUELQUES DÉFIS DE RECHERCHE CONCERNANT LA COMPRÉHENSION DES CAPACITÉS DES AGENTS CONVERSATIONNELS PAR LES UTILISATEURS

- Quelles données les chatbots enregistrent-t-ils ? Sont-elles anonymisées ?
- Comment peut-on mener des audits du comportement des chatbots (mesure automatique ou/et évaluation humaine) ?
- Les répliques sélectionnées par les chatbots sont-elles explicables ? Les chatbots peuvent-ils les rendre eux-mêmes plus compréhensibles ?
- Quels paramètres du profil de son interlocuteur les chatbots calculent-ils ? Les humains en sont-ils conscients ?
- L'idée que l'utilisateur se fait de la stratégie du chatbot correspond-elle à la stratégie réelle mise en place dans le chatbot ?

[68] "Google Assistant", "Google Home", "Apple Siri", "Amazon Alexa" et "Amazon Echo", "Yandex Alisa", "Mail.ru Marusia", "Baidu DuerOS", "Xiaomi XiaoAI", "Tencent Xiaowei", "Samsung Bixbi", "Orange Djingo", etc.

[69] A. Turing, "Computing Machinery and Intelligence", Mind 59(236) 433–460, 1950.

[70] J. Weizenbaum, "ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine", Communications of the Association for Computing Machinery 9, 36-45, 1966.

[71] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum, "The Design and Implementation of Xiaoice, an Empathetic Social Chatbot", Computational Linguistics 46(1), 53-93, 2020.

QUESTIONS ÉTHIQUES

Le langage est un élément constitutif de l'identité de l'être humain et le fondement de sa vie en société. Les agents conversationnels sont ainsi naturellement comparés à un être humain, que leur interlocuteur soit informé, ou non, de leur caractère artificiel. Cet aspect naturel du dialogue est susceptible d'influer sur l'être humain : c'est le problème fondamental de l'éthique des chatbots. Leur déploiement étant un phénomène récent, on ne dispose pas de données expérimentales suffisantes pour évaluer leurs effets sur l'être humain à long terme.

Depuis peu, les performances de la reconnaissance de la parole permettent l'utilisation des interfaces vocales. Outre le dialogue langagier, la voix porte des informations de diverses natures, par exemple sur l'âge, le sexe, la corpulence, la langue maternelle, l'accent, les lieux de vie, le milieu socio-culturel, l'éducation, l'état de santé, la compréhension ou les émotions de la personne qui parle. De nombreuses questions éthiques sont liées à ces aspects de la vie humaine.

À l'instar des systèmes techniques en général et des systèmes autonomes en particulier (par exemple, la reconnaissance automatique d'images ou la conduite autonome des véhicules), les agents conversationnels doivent répondre à un grand nombre d'exigences en termes de sécurité, transparence, traçabilité, utilité, protection de la vie privée, etc. Les systèmes de chaque type mettent ces propriétés en œuvre en fonction du contexte spécifique de leur utilisation. Dans tous les cas, il s'agit de contraintes de premier plan pour le concepteur comme pour l'utilisateur.

Certains agents conversationnels provoquent des tensions éthiques nouvelles, par exemple liées à l'impossibilité d'expliquer en langue naturelle la chaîne des décisions aboutissant à telle ou telle recommandation médicale. Des préconisations sont formulées à cet égard dans l'avis de la CERNA sur les questions éthiques de la recherche en apprentissage machine[72].

[72] http://cerna-ethics-allistene.org/Publications%2bCERNA/apprentissage/index.html

CONSULTATION

I.LES FACTEURS ÉTHIQUES DANS L'UTILISATION DES CHATBOTS

1) CONFUSION DE STATUT

Plusieurs facteurs contribuent à faire confondre un agent conversationnel avec un être humain. Un effacement des distinctions de statut peut advenir comme une brève illusion ou, au contraire, il peut persister tout au long d'un dialogue. Il peut également être volontaire ou spontané, avoir des conséquences psychologiques ou juridiques, donner lieu à des manipulations plus ou moins graves. Cette confusion de statut a pour cause un phénomène plus général.

Quelle que soit la nature de son interlocuteur, l'être humain projette sur lui spontanément des traits humains : pensée, volonté, désir, conscience, représentation interne du monde. Ce comportement est qualifié d'« anthropomorphisme ». L'interlocuteur apparaît alors comme un individu autonome doté de pensée propre, qu'il exprime à travers sa parole.

A ce jour, seule une loi de l'État de Californie[73] impose explicitement de mentionner l'existence d'une interaction avec un chatbot lorsque cette interaction entend inciter à l'achat ou vendre des produits ou services dans le cadre d'une transaction commerciale ou influencer le vote dans un cadre électoral. Il n'existe pas d'équivalent à cette disposition dans le droit français ou européen même si une réflexion est désormais engagée sur ce point[74].

1.1 Faut-il informer l'utilisateur de la nature de son interlocuteur (être humain ou machine) ? Si oui, quelles informations sur le chatbot faut-il communiquer à l'utilisateur (finalités, corpus d'entraînement, nom du concepteur, etc.) ?

1.2 Pensez-vous qu'en Europe, il faudrait adopter un cadre législatif comparable à celui de l'État de Californie ?

1.3 Remarque libre :

2) ATTRIBUTION DE NOM PROPRE

Souvent, l'être humain donne à un agent conversationnel un nom, comme par exemple les enfants le font avec leurs poupées.

Parfois, l'attribution du nom est voulue par le concepteur : s'adresser à la machine par un nom peut aider à mieux réaliser sa fonction, par exemple dans les secteurs d'assistance aux personnes ou de divertissement. Dans ces cas, l'utilisation du nom renforce la réaction émotionnelle de l'utilisateur.

Actuellement, ce recours au nom et à la réaction émotionnelle sert encore souvent à masquer le manque de performances sémantiques et contextuelles des agents conversationnels. Attribuer un nom à la machine relève de la dynamique de projection, c'est-à-dire d'anthropomorphisation de cette machine. Or, lorsque l'agent conversationnel lui-même emploie son « nom » dans un dialogue, se pose alors la question de l'autoréférence : à qui ou quoi exactement renvoie ce nom ?

2.1 L'utilisateur devrait-il pouvoir choisir le nom et le genre du nom (masculin, féminin, neutre) porté par un chatbot ou ce choix relève-t-il du concepteur ?

2.2 Un chatbot pourrait-il ou devrait-il se voir attribuer un nom humain (par exemple « Sophia »), un nom non-humain (par exemple « R2D2 ») ou bien aucun nom ?

2.3 Remarque libre :

3) MALMENER LES CHATBOTS

La projection des qualités humaines sur les agents conversationnels est un phénomène courant et important. En particulier, les utilisateurs pourraient maltraiter un agent conversationnel.

Tandis que votre chatbot vous rappelle les gestes barrière pendant une épidémie, vous pourriez réagir en l'insultant ou en lui ordonnant de se taire. En outre, cela pourrait avoir une incidence sur les enfants qui entendent cet échange.

Les assistants vocaux généralistes (Siri, etc.) se font parfois insulter par les utilisateurs. Dans ce cas, ils répondent selon des stratégies prédéterminées par leurs concepteurs.

3.1 Insulter un chatbot dans une conversation est-ce un acte moralement répréhensible ? Pensez-vous qu'il est admissible de se servir du chatbot comme un souffle-douleur ?

3.2 Un chatbot insulté par son interlocuteur devrait-il pouvoir répondre à l'utilisateur en l'insultant au retour ?

3.3 Si un chatbot répondant à un nom féminin voire ayant une voix féminine est malmené, y voyez-vous un geste de maltraitance envers les femmes ? La même question se pose pour les noms masculins.

3.4 Remarque libre :

4) CONFIANCE DANS LES CHATBOTS

Une certaine confiance de l'utilisateur envers les finalités du chatbot est nécessaire pour la réalisation des tâches fonctionnelles du chatbot.

La confiance n'est pas seulement un phénomène psychologique émergent mais relève d'un effort technique : les concepteurs des agents conversationnels cherchent à l'établir et à la maintenir, mais pourraient également se poser la question d'éviter qu'elle soit accordée au chatbot de manière irréfléchie.

L'évaluation du niveau de confiance des utilisateurs envers les comportements et performances du chatbot est un important sujet de recherche.

4.1 Si la réponse « je ne sais pas » d'un chatbot vient en conflit avec la préservation de la confiance de l'utilisateur, par exemple dans le cas d'un service après-vente, faut-il privilégier la confiance en modifiant la réponse ?

4.2 Afin de gagner la confiance, le chatbot peut-il se présenter comme un.e « assistant.e / conseiller.ère / ami.e » de l'utilisateur ?

4.3 Remarque libre :

[73] https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=BPC&division=7.&title=&part=3.&chapter=6.&article

[74] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

5) LES CONFLITS DES CHATBOTS

Si la plupart des systèmes de dialogue sont conçus pour une tâche spécifique, de nombreux autres sont des agents conversationnels généralistes. Leur interaction avec l'être humain peut participer à un conflit. On se pose alors la question du rôle de l'agent conversationnel dans ce conflit et du jugement qui va tomber sur lui. Par exemple, un chatbot pourrait donner de fâcheux conseils à son utilisateur, lui mentir, ou encore se comporter en délateur en appelant la police s'il détecte à tort ou à raison une menace.

Les recherches actuelles portent sur le développement et l'utilisation des systèmes capables de s'adapter aux utilisateurs, à leurs desiderata, intentions et croyances en leur répondant comme le ferait un proche. Ces réponses adaptées voire « intelligentes » en réaction aux questions ou comportements des humains ne peuvent qu'engendrer chez les utilisateurs des croyances sur des « compétences » ou des supposés « états d'esprit » de la machine. L'humain s'adapte ainsi aux agents conversationnels avec lesquels il dialogue, soit en s'en méfiant, soit au contraire en leur donnant un certain « crédit de vérité ». En s'appuyant sur son « crédit de vérité », un chatbot pourrait proférer un mensonge.

La tension émerge lorsque le chatbot, par exemple, répond à une question de l'utilisateur relative à sa santé. Un médecin peut le cas échéant cacher toute la vérité à son patient dans le souci du bien-être de celui-ci.

5.1 Le mensonge proféré par un chatbot est-il plus ou moins acceptable que le mensonge humain ? La réponse dépend-elle du contexte (assistant vocal, éducation, psychothérapie, recrutement, etc.) ?

5.2 Si les chatbots peuvent mentir aux utilisateurs, qui et comment devrait décider des buts admissibles et des limites de tels comportements ?

5.3 Remarque libre :

6) LA MANIPULATION (NUDGE) DES CHATBOTS

Prix Nobel d'économie, l'Américain Richard Thaler a mis en lumière le concept de nudge, qui consiste à inciter les individus à changer de comportement sans les contraindre, par la seule utilisation de leurs biais cognitifs. Dans le cas des chatbots, les nudges sont définis comme des suggestions ou manipulations, manifestes ou cachées, conçues pour influencer le comportement ou les émotions d'un utilisateur.

Les agents conversationnels pourraient ainsi devenir un moyen d'influence des individus à des fins mercantiles ou politiques. Mais le nudge est aussi souvent mis en œuvre pour surveiller notre santé ou pour améliorer notre bien-être (faire plus d'exercice physique, consommer moins d'alcool, arrêter de fumer, etc.).

6.1 Tous les nudges sont-ils permis ? Comment peut-on distinguer les bons des mauvais nudges ?

6.2 Le concept de consentement libre et éclairé dans le cadre d'un agent conversationnel capables de « nudge » a-t-il encore un sens ?

6.3 Remarque libre :

7) LES CHATBOTS ET LE LIBRE CHOIX

Lors d'un dialogue, les chatbots évaluent plusieurs réponses possibles pour en donner une seule. Dans le cas des systèmes de recommandation, ce choix unique pourrait limiter la liberté de l'utilisateur de choisir de manière autonome, en dérobant à sa vue toute la palette d'options disponibles. Cela génère en outre un risque d'enfermement (filter bubble), problème renforcé par le faible niveau de paramétrage proposé par les systèmes commercialisés actuellement.

Par exemple, à la demande de commander une pizza, le chatbot propose de commander chez un restaurateur particulier. Cela peut être le fournisseur plus proche géographiquement, le mieux noté sur un site donné ou encore celui qui possède un accord commercial avec le concepteur du chatbot. Or il propose un choix unique tandis qu'il existe au voisinage quinze pizzerias qui proposent le service demandé. Ce choix unique peut poser un problème éthique lié à la liberté et à la discrimination.

7.1 Dans l'exemple cité, souhaiteriez-vous que le chatbot explicite tous les choix ou plusieurs choix ?

7.2 Pensez-vous qu'une information transparente de l'utilisateur sur les critères de choix des recommandations par le chatbot soit une solution satisfaisante aux problèmes éthiques du libre choix et de la discrimination ?

7.3 Remarque libre :

8) LES ÉMOTIONS DES CHATBOTS

Les émotions sont souvent mélangées dans la vie de tous les jours. En conséquence, la détection et l'identification des émotions des utilisateurs dépendent d'un grand nombre de facteurs contextuels, culturels et idiosyncrasiques. L'informatique émotionnelle comprend trois grands domaines : détecter les émotions des humains, raisonner sur ces informations pour modifier la stratégie du dialogue et générer une expressivité émotionnelle par le langage ou le comportement non-verbal du chatbot.

Par exemple, ayant reconnu que l'utilisateur est stressé, un agent conversationnel peut simuler l'empathie et exprimer la compréhension de l'état de l'utilisateur.

8.1 Est-il souhaitable de construire des chatbots qui détectent les émotions des êtres humains ? Précisez la réponse selon le contexte d'utilisation.

8.2 Est-il souhaitable de construire des chatbots qui simulent des émotions des êtres humains ? Précisez la réponse selon le contexte d'utilisation.

8.3 Remarque libre :

9) LES CHATBOTS ET LES PERSONNES VULNÉRABLES

Un chatbot peut occuper toute l'attention d'une personne vulnérable en remplaçant, comme dans le cas des enfants autistes, le difficile contact avec les personnes humaines. Ce phénomène provoque souvent des jugements polarisés : d'un côté, le bien-être de la personne peut être amélioré ; de l'autre, il l'est au dépens de sa socialisation humaine « standard ».

Par exemple, un enfant autiste pourrait préférer l'interaction très nourrie et prolongée avec un chatbot à celle avec un parent ou un pédagogue. Un jeune enfant pourrait apprendre et imiter les comportements émotionnels de la machine au lieu de ceux des humains. Une personne âgée pourrait vouloir faire le deuil de son chatbot ou l'enterrer si elle lui est très attachée et qu'il ne fonctionne plus.

9.1 Quelles finalités de l'interaction entre un chatbot et une personne vulnérable (surveillance, éducation, accompagnement, divertissement) sont acceptables ? La réponse dépend-elle de l'âge de la personne (enfant, personne âgée) ou de son statut (patient, personne en convalescence) ?

9.2 Les utilisateurs, notamment les personnes vulnérables, sont susceptibles de s'attacher profondément à des chatbots, ce qui peut entraîner une modification durable de leur façon de vivre ou de leurs interactions sociales. Ce phénomène est-il inquiétant ? Pourquoi ?

9.2 Remarque libre :

10) LES CHATBOTS ET LA MÉMOIRE DES MORTS

Si le droit à la vie privée s'éteint à la mort de la personne, l'utilisation post-mortem de ses données, par exemple de sa voix, par un chatbot pour faire « revivre » cette personne peut néanmoins poser problème quant à l'atteinte possible au principe de respect de la dignité de la personne humaine.

Un journaliste américain est parvenu à créer un chatbot, le « dadbot », à partir des souvenirs qu'il avait de son père[75]. Il échange avec ce chatbot « comme si » il s'agissait d'un échange avec son père.

10.1 Pensez-vous que les chatbots sont un moyen envisageable pour faire « vivre » la mémoire ou la manière de s'exprimer propres à une personne décédée ? De tels usages porteraient-ils atteinte au principe de respect de la dignité de la personne humaine ?

10.2 Quelle évolution du concept de mort envisagez-vous en tenant compte des possibilités offertes par les chatbots ?

10.3 Remarque libre :

11) SURVEILLANCE PAR LES CHATBOTS

Si certains chatbots font partie des systèmes exclusivement consacrés à l'interaction humain-machine, d'autres fonctionnent dans des environnements partagés. Les chatbots capables d'enregistrer la voix pourraient ainsi surveiller les interactions autour d'eux, que celles-ci soient humaines ou avec d'autres chatbots. Cette capacité implique des enjeux éthiques et juridiques liées à la protection de la vie privée, à l'exploitation des données personnelles sans consentement, au risque de violation du secret personnel ou professionnel ainsi qu'à l'introduction de failles de sécurité. La divulgation par les chatbots des contenus enregistrés à l'insu des personnes peut s'apparenter à la délation.

Par exemple, en cas d'écart à la diète que le médecin a imposée à un patient, le chatbot l'en informe, voire se met en contact l'organisme de soins de santé.

Autre exemple, un chatbot peut « tenir compagnie » des personnes vulnérables ou âgées en surveillant leur comportement.

11.1 Dans les exemples cités, pensez-vous que le comportement du chatbot est justifié ? Comment, dans ce cas, l'utilisateur peut-il exprimer son consentement ? Qu'en est-il si les chatbots sont placés dans des espaces partagés ?

11.2 Donnez d'autres exemples de situation dans laquelle la surveillance par un chatbot vous paraît justifiée.

11.3 Si un chatbot est insulté par son utilisateur, cette information doit-elle être communiquée par le chatbot à une tierce partie, par exemple son concepteur ?

11.4 Remarque libre :

12) LES CHATBOTS ET LE TRAVAIL

Les chatbots présenteront des opportunités et des risques pour les entreprises selon les contextes de leur utilisation (évaluation, recrutement, divertissement, etc.). L'introduction d'agents conversationnels dans les équipes peut induire des effets organisationnels selon les secteurs industriels, notamment du point de vue de la charge informationnelle et émotionnelle, de la temporalité du travail, du sentiment de cohésion ou d'isolement des travailleurs, des effets des chatbots sur le moral des employés ainsi que les problèmes d'égalité et de reconnaissance au mérite au sein des entreprises.

Par exemple, dans le secteur médical, l'aide à l'action humaine (médecins psychiatres, médecins généralistes, infirmiers, agents des centres d'appel d'urgence, etc.) par des chatbots pourrait provoquer des effets sur la profession dans sa totalité ainsi que sur le bien-être des patients et des personnels soignants et sur la relation entre eux.

12.1 Existe-t-il des métiers ou des pratiques humaines dans lesquels le recours aux chatbots devrait être encouragé ou prohibé ?

12.2 Comment et à quelle échelle temporelle envisagez-vous l'évolution des métiers à la suite de l'introduction des chatbots ? Précisez votre réponse selon un ou plusieurs cas d'usage.

12.3 Par quels moyens (législatif, code de bonne conduite, etc.) le recours aux chatbots devrait-il être encadré ?

12.4 Remarque libre :

13) EFFETS À LONG TERME SUR LE LANGAGE

À moyen et long termes, l'utilisation des chatbots peut avoir une incidence durable sur le langage humain et peut-être également sur les habitudes de vie.

Par exemple, si les chatbots répondent par des phrases courtes, linguistiquement pauvres, sans politesse aucune, les humains risquent d'imiter ces tics langagiers lorsqu'ils s'adressent à d'autres humains.

13.1 Comment envisagez-vous l'évolution du langage sous l'influence des chatbots ? Cette influence peut-elle être jugée comme bonne ou mauvaise ?

13.2 Quelle échelle temporelle peut-on envisager pour cette évolution ?

13.3 Remarque libre :

[75] James Vlahos. Talk to me, Amazon, Google, Apple, and the Race for Voice-Controlled AI. Random House, 2019.

II. LES FACTEURS ÉTHIQUES DANS LA CONCEPTION DES CHATBOTS

14) PROBLÈME DE SPÉCIFICATION

Les lois et les règles de conduite dans la société sont formulées dans une langue naturelle. Leur traduction dans un langage informatique exige une « spécification » : définition de tous les termes dans un cadre formel. Souvent, la spécification complète est impossible : par exemple, le terme « humain » peut inclure des humains qui seraient facilement identifiables par un système informatique apprenant, mais aussi des humains que le système ne parviendra pas à identifier comme tels car absents des données d'apprentissage. Quels que soient la base d'apprentissage et l'algorithme déployé, les erreurs d'identification sont inévitables : par nature, la langue humaine admet la multiplicité des significations.

Pour les chatbots, le problème de spécification se traduit, par exemple, par la difficulté de distinguer systématiquement et sans erreur, l'usage ironique ou satirique d'un concept ou d'une expression de son usage indicatif standard.

14.1 Quelles erreurs commises par les chatbots seraient acceptables et lesquelles ne le seraient pas ? Précisez la réponse selon le contexte (santé, éducation, divertissement, service après-vente, etc.).

14.2 Si un chatbot n'est pas capable de trouver une réponse, doit-il le dire explicitement ?

14.3 Quelles conséquences sur le comportement des utilisateurs la réponse « je ne sais pas », fréquemment donnée par les assistants vocaux actuels, entraîne-t-elle ? Si vous avez vécu cette expérience, décrivez-la.

14.4 Remarque libre :

15) LES MÉTRIQUES ET LES FONCTIONS D'ÉVALUATION

Dans un agent conversationnel, les fins recherchées par le concepteur donnent lieu à la définition d'une métrique ou d'une fonction d'évaluation, qui quantifie la mesure de « bonne réponse » ou « réplique adéquate » pour le système. Cette métrique est encodée au préalable. La métrique d'un chatbot peut aussi tenir compte des facteurs émergents, qui apparaissent pendant la conversation, par ailleurs susceptibles de causer des ruptures dans la compréhension humaine du comportement du système. Souvent, cette qualité du dialogue est mesurée par le degré d'engagement de l'utilisateur, c'est-à-dire sa volonté à poursuivre le dialogue avec le chatbot. La métrique d'engagement utilise la longueur des échanges comme des marqueurs paralinguistiques (rire, sourire, hésitation, hochement de tête, etc.) de satisfaction ou d'intérêt ; or, dans l'état actuel des recherches, elle tient rarement compte du contenu sémantique des échanges. Cela peut défavoriser ceux qui ne comprennent pas le procédé d'évaluation de l'agent conversationnel et en outre donner lieu à des comportements manipulateurs de la part des utilisateurs.

En avril 2016, le chatbot Tay de Microsoft, qui avait la capacité d'apprendre en continu à partir de ses interactions avec les internautes, avait appris à tenir des propos racistes. Tay a été rapidement retiré par Microsoft.

Malgré cette expérience, DeepCom, un autre chatbot développé par Microsoft China en 2019 afin de commenter des nouvelles sur les réseaux sociaux, a été reconnu par ses concepteurs eux-mêmes comme étant susceptible de générer des contenus biaisés, (par exemple, discriminants) voire de la propagande, à la suite de fortes réactions dans la communauté de recherche[76]. La première version de la publication postulait : « Compte tenu de la prévalence des articles de presse en ligne avec commentaires, il est très intéressant de mettre en place un système de commentaire automatique des nouvelles avec des approches construites à partir des données ». Dans la version révisée, les auteurs affirment : « Il existe un risque que des personnes et des organisations utilisent ces techniques à grande échelle pour simuler des commentaires provenant de personnes à des fins de manipulation ou de persuasion politique ».

15.1 Faudrait-il que l'utilisateur soit informé du fait que la stratégie de dialogue d'un chatbot puisse être adaptée au cours de la conversation ?

15.2 Comme expliqué plus haut, l'utilisateur peut manipuler la métrique d'un chatbot à ses propres fins. S'il le fait, le concepteur partage-t-il l'éventuelle responsabilité pour les résultats de cette manipulation ou devrait-il en être dédouané?

15.3 Avez-vous vécu des exemples personnels liés, selon votre interprétation, aux métriques particulières des chatbots ?

15.4 Remarque libre :

16) LES FINALITÉS DE L'AGENT CONVERSATIONNEL

Les finalités d'un chatbot, c'est-à-dire les buts qui lui sont assignés, sont définies par ses concepteurs, et le chatbot cherche à les satisfaire dès sa mise en marche. Si cela ne pose pas de problèmes excessifs pour les chatbots dédiés à une ou plusieurs tâches connues au préalable, la spécification des finalités peut s'avérer complexe pour un chatbot généraliste car elles ne sont pas toutes énumérables au moment de la conception.

Ces finalités peuvent être très diverses : des systèmes après-vente aident à réparer des produits défectueux, des conseillers médicaux cherchent à améliorer l'état du patient, des services d'aide au recrutement, etc.

D'autres systèmes possèdent des finalités plus vagues : certains chatbots sont conçus afin de converser librement avec l'utilisateur sur tous les sujets. Que la perception des finalités ou le jugement que l'on porte sur elles puissent évoluer, cela ne supprime guère cette distinction fondamentale entre un agent conversationnel et un humain qui peut agir sans finalité prédéterminée et peut ne pas rendre sa finalité transparente aux autres.

16.1 Doit-on révéler la finalité d'un chatbot à l'utilisateur ? Si oui, à quel moment et sous quelle forme ? Si non, pourquoi ?

16.2 Devrait-on accepter qu'un chatbot capable d'apprentissage en interaction (par exemple, un agent conversationnel généraliste) puisse être dirigé vers une finalité particulière à travers une influence intentionnelle ou involontaire de la part des utilisateurs (par exemple, inciter la personne à faire un don ou à acheter un produit particulier) ? Précisez la réponse selon le contexte (santé, éducation, divertissement).

16.3 Remarque libre :

[76] https://www.vice.com/en/article/d3a4mk/microsoft-used-machine-learning-to-make-a-bot-that-comments-on-news-articles-for-some-reason

17) LES BIAIS D'APPRENTISSAGE

Un système apprend à partir de données sélectionnées par un « entraîneur » (agent humain responsable de leur sélection). L'existence de biais dans les données d'apprentissage est une source majeure des conflits éthiques, notamment à travers la discrimination ethniques, culturelles ou encore de genre.

Par exemple, des données de parole enregistrées peut contenir uniquement des voix d'adultes alors que le système est censé interagir aussi avec les enfants, ou un corpus de textes peut utiliser statistiquement plus fréquemment des pronoms de genre féminin que ceux de genre masculin.

Le système reproduira alors ces biais issus d'un corpus d'apprentissage, sauf s'il est équipé d'outils spécialement conçus dans le but de les corriger, ce qui présuppose déjà la connaissance des biais possibles. Or, certains biais pourraient ne pas être connus à l'avance.

17.1 Considérez-vous qu'un agent conversationnel devrait être sans biais ? Est-ce possible ? Précisez la réponse selon le contexte (santé, recrutement, service après-vente, éducation, sécurité, assistant vocal domestique).

17.2 Pensez-vous que les chatbots devraient imiter les biais humains ou les corriger ?

17.3 Remarque libre :

18) INSTABILITÉ DE L'APPRENTISSAGE

Des erreurs sont inévitables lorsqu'un système apprenant classifie une donnée qui ne ressemble pas, ou qui ressemble faussement, à celles contenues dans le corpus utilisé pendant son apprentissage. Dans le cas des agents conversationnels, cela recouvre les homophones, homographes, homonymes ou autres exemples d'ambiguïté linguistique.

Un cas simple est celui des erreurs d'orthographe : le comportement du chatbot dans ce cas diffère totalement de celui de l'être humain. Par exemple, l'utilisateur humain reconnaît un mot même s'il contient plusieurs erreurs, tandis qu'à cause de l'instabilité, un algorithme cesse de reconnaître correctement un mot contenant une ou deux fautes d'orthographe.

18.1 L'apprentissage des chatbots étant instable, il induit des erreurs parfois évidentes. Êtes-vous prêt à tolérer ces erreurs davantage que les erreurs humaines ? Précisez la réponse selon le contexte.

18.2 Les erreurs des chatbots provoquent-elles des sentiments ou des réactions différentes par rapport aux erreurs humaines ? Lesquelles ?

18.3 Remarque libre :

19) EXPLICABILITÉ ET TRANSPARENCE

La transparence d'un système signifie que son fonctionnement n'est pas opaque ou incompréhensible pour l'homme. Elle s'appuie notamment sur la traçabilité des répliques sélectionnées par un agent conversationnel. L'explicabilité signifie qu'un utilisateur peut appréhender le comportement du chatbot. Les problèmes de transparence et d'explicabilité sont provoqués par différents facteurs, notamment par le fait que, contrairement à l'être humain, un système informatique ne comprend pas le sens des phrases qu'il génère ou qu'il perçoit.

Ainsi, un chatbot, qui n'a pas de représentation du monde, est susceptible de formuler des phrases qui ne correspondent à aucune réalité (« lait noir »), de répondre sans tenir compte du contexte (« Comment vas-tu ? » - « Il fait beau ») ou d'employer un lexique désagréable ou prohibé.

Les effets immédiats sur l'utilisateur provoqués par un tel dialogue peuvent être importants (réaction émotionnelle forte, rupture dans la compréhension, abandon du dialogue ou débranchement du système). La question de responsabilité se pose alors à l'égard des concepteurs et des entraîneurs des agents conversationnels. La dimension esthétique (certaines paroles peuvent être étranges mais belles) suffit-elle à dédouaner le chatbot du besoin d'imiter toujours la parole humaine ?

19.1 À quelle réaction peut-on s'attendre de la part d'un utilisateur en situation de rupture de compréhension dans un dialogue avec le chatbot ? Précisez la réponse selon les finalités de celui-ci et le contexte (par exemple, santé, assistant vocal généraliste, divertissement, recrutement).

19.2 Lorsque l'utilisateur donne spontanément un sens à des répliques peu compréhensibles du chatbot, ce phénomène relève-t-il d'une attitude ludique ou pose-t-il un problème éthique ?

19.3 Remarque libre :

20) IMPOSSIBILITÉ D'ÉVALUATION RIGOUREUSE

Un agent conversationnel fournit une réponse en appliquant des stratégies de dialogue qui dépendent de l'interprétation. Les modèles les plus avancés utilisent de grands corpus de données pour apprendre.

L'évaluation de ce système de dialogue, par essence dynamique, est difficile au moins sur deux plans : a) la prédiction des entrées générées par l'utilisateur n'est souvent pas possible; b) les aléas de l'apprentissage contribuent à la difficulté de reproduire le comportement du système.

Or, l'incertitude théorique et pratique va de pair avec les techniques d'apprentissage qui procurent aux systèmes leur grande efficacité.

20.1 Est-ce acceptable qu'un chatbot profère des phrases « incongrues », qu'aucun être humain n'a jamais utilisées, ce qui serait susceptible d'influencer son interlocuteur ?

20.2 Un chatbot devrait-il se limiter à un ensemble prédéterminé de phrases ou, à l'inverse, en générer librement ? Précisez la réponse selon le contexte (divertissement, service après-vente, éducation, assistant vocal généraliste).

20.3 Remarque libre :

MERCI BEAUCOUP POUR VOTRE CONTRIBUTION !

L'ENVOI SE FAIT À L'ADRESSE CNPEN-CONSULTATION-CHATBOTS@CCNE.FR

ANNEXE 3 : COMPOSITION DU GROUPE DE TRAVAIL

Laurence Devillers et Alexei Grinbaum, co-rapporteurs
Gilles Adda
Raja Chatila
Caroline Martin
Serena Villata
Célia Zolynski

Ont également contribué :
Eric Germain
Christophe Lazaro
Félicien Vallet (CNIL)
Camille Darche (secrétaire)

ANNEXE 4 : MODE DE TRAVAIL

Auditions :

- Pr Pierre Philip (PU PH), Université de Bordeaux. L'USR SANPSY (sommeil, addiction, neuropsychiatrie) USR3413, CNRS - Université Bordeaux 2 - 24/02/20
- Julia Velkovska, sociologue à l'EHESS et chez Orange - 20/04/20
- Mickaël Cabrol, (www.easyrecrue.com) - Easyrecrue CEO - Arthur Guillon, Senior Machine Learning Engineer - Léo Hemamou, doctorant, thèse en détection automatique de signaux sociaux - 15/05/20

Consultations : https://www.ccne-ethique.fr/en/actualites/cnpen-ethical-issues-conversational-agents
https://www.ccne-ethique.fr/fr/actualites/cnpen-les-enjeux-ethiques-des-agents-conversationnels

- Le comité a lancé une consultation sur des questions éthiques propres aux chatbots de juin à octobre 2020. Cette consultation identifie différents enjeux et propose des scénarios d'utilisation de plusieurs types de chatbots.
- L'appel visait à permettre une expression des parties prenantes et du public sur les enjeux éthiques liés aux chatbots. Chaque contributeur était invité à répondre à l'ensemble des questions posées. Il était spécifié que les propos des contributeurs seraient anonymisés et ne seraient pas cités nommément dans l'avis.
- Le comité a recueilli l'avis d'une centaine de répondants (personnes physiques, institutions publiques et privées).
- Le comité n'a pas souhaité analyser les résultats de manière quantitative ; trois séances de travail ont été consacrées à l'étude de l'ensemble des réponses, dans le but de nourrir la réflexion collective.