



AVIS n°2024-46

**« LE PHÉNOMÈNE D'ATTACHEMENT AUX ROBOTS DITS « SOCIAUX ».
POUR UNE VIGILANCE DE LA RECHERCHE SCIENTIFIQUE »**

Approbation le 1^{er} juillet 2024

MEMBRES DU GROUPE DE TRAVAIL :

Catherine Pelachaud, membre du COMETS, rapporteure

Patrice Debré, membre du COMETS, rapporteur

Christine Noiville, membre du COMETS

Raja Chatila, professeur émérite d'Intelligence artificielle, de robotique et d'éthique à Sorbonne Université

Jean-Gabriel Ganascia, professeur émérite d'informatique et d'intelligence artificielle à Sorbonne Université



SOMMMAIRE	2
I. RESUMÉ	3
II. AUTO-SAISINE	5
III. ANALYSE	6
1. Les robots sociaux : une place croissante dans nos vies quotidiennes	6
2. Les effets individuels et collectifs de l'attachement aux robots sociaux	9
3. Pour une contribution réfléchie et responsable de la recherche publique au développement des robots sociaux	12
IV. RECOMMANDATIONS	13
V. PERSONNALITÉS CONSULTÉES	15
VI. BIBLIOGRAPHIE	15
VII. LISTE DES ACRONYMES	18

I. RESUMÉ

Le COMETS a souhaité se saisir d'une question qui lui semble avoir été insuffisamment prise en compte par le monde de la recherche publique, celle des robots dits « sociaux » et, plus précisément, des impacts cognitifs et psychologiques liés à l'utilisation croissante de ces outils dans la vie quotidienne.

Les chatbots, agents conversationnels et autres robots de compagnie programmés avec des techniques d'intelligence artificielle et embarqués dans toute une série d'objets connectés – ordinateurs, téléphones, montres, automobiles – font désormais partie intégrante de l'environnement quotidien. Or une quantité croissante d'entre eux sont conçus dans une logique affective, pour faire office de compagnon, de confident, d'ami, de coach de santé ou de bien-être, ou pour remplacer un proche défunt (*deadbot*). Ils utilisent souvent des attributs propres aux humains (langage, apparence, attitude), sont capables d'interagir avec leur utilisateur sur le modèle des humains (voix, intonation, gestuelle, expressions faciales) et, à l'aide de capteurs audio ou de caméras, prétendent détecter ses émotions (vous êtes triste ? Vous avez l'air angoissé(e) !) et en simuler eux-mêmes (en pleurant avec l'utilisateur, en riant avec lui, en le félicitant, etc.). L'utilisateur tend alors à attribuer à la machine des capacités humaines (intelligence, conscience, bienveillance, empathie), à se projeter dans une interaction affective avec elle, à développer l'illusion que se noue un lien intime et de confiance entre elle et lui, voire à s'y attacher.

Tout en étant conscient que ce phénomène peut engendrer certains bénéfices, le COMETS est préoccupé par les impacts individuels et collectifs qui peuvent en résulter, notamment en termes de dépendance affective, d'addiction, d'emprise, de manipulation, de manque d'interactions avec autrui voire de désocialisation, etc.

Il adhère aux recommandations déjà formulées dans divers cadres (CNPEN, CERNA, littérature juridique et éthique...) en direction des industriels et des ingénieurs concepteurs de robots sociaux, d'une part, et des pouvoirs publics d'autre part. Il s'agit notamment de développer de manière réfléchie et responsable ces dispositifs dès leur conception pour éviter toute manipulation des utilisateurs, d'informer toute personne qui communique avec un robot du fait qu'elle dialogue avec une machine, d'éviter la possibilité technique de manipulations malveillantes, de menaces proférées par le robot, d'exploitation des émotions contraire à l'intégrité et à l'autonomie des personnes, etc.

Le COMETS estime toutefois nécessaire d'appeler spécifiquement à la vigilance les chercheurs, les sociétés savantes et les institutions de la recherche publique, et cela à deux titres.

D'une part, que ce soit au CNRS, à l'INRIA, au CEA, dans diverses universités, un certain nombre de travaux en informatique, robotique, sciences du comportement, traitement du langage, contribuent à conforter le phénomène d'attachement des utilisateurs aux robots sociaux, sans réflexion suffisante sur les finalités recherchées et les effets. S'il est louable de chercher à améliorer l'interface humain-machine pour un meilleur « engagement » des utilisateurs, il est nécessaire de questionner davantage les inconvénients liés à l'anthropomorphisation des robots (en termes d'apparence et de comportements) et aux impacts émotionnels et psychologiques qui l'accompagnent.

D'autre part, la recherche publique a un rôle de premier plan à jouer pour suivre et mesurer les conséquences à long terme de l'utilisation des robots sociaux. Maintenant que ces derniers sont utilisés à grande échelle, il s'agit d'en mesurer l'incidence sur la cognition, le psychisme, le comportement des utilisateurs, le rapport de ces derniers à autrui et au monde ; de construire ainsi un socle de savoirs nécessaires pour faire face aux enjeux liés à l'usage de ces outils et pour en assurer une utilisation responsable et libre.

Le COMETS recommande donc que la recherche publique (chercheurs en informatique et robotique, sociétés savantes, organismes de recherche) :

1. développe des formations aux enjeux éthiques (dans les cursus scientifiques et techniques et pour les personnels de recherche concernés), se familiarise davantage avec la littérature internationale consacrée à ces questions et en débata collectivement ;
2. s'interroge sur les finalités de la recherche, des applications et des choix de conception, comme sur les avantages et les inconvénients à donner aux robots une forme ou un comportement humanoïde ou des capacités à capter et à simuler les émotions ;
3. mène, dans des situations et contextes réalistes, des études scientifiques de long terme et à grande échelle sur : - les relations que les utilisateurs nouent avec « leurs » robots sociaux et sur les incidences en termes de cognition, de psychisme, d'attachement, d'autonomie d'actions et de décisions ; - les effets du déploiement des robots sociaux sur les relations *entre* humains ;
4. renforce à cet effet les recherches interdisciplinaires et indépendantes associant aux travaux en informatique, en robotique, en sciences du comportement, du traitement du langage, etc., des recherches en psychologie, neurosciences, linguistique, sociologie, droit, éthique, philosophie, anthropologie ;
5. dans le cadre d'un observatoire, collecte à grande échelle et à long terme les données relatives à l'utilisation de robots sociaux, la manière dont les utilisateurs se les approprient, les impacts sur leurs états émotionnels et leurs décisions ; l'objectif est d'alimenter la recherche scientifique et, à plus long terme, d'éclairer les utilisateurs et les décideurs sur les conditions d'un développement et d'un usage libre et responsable de ces dispositifs.

II. AUTO-SAISINE

Le déploiement à grande échelle d'objets programmés avec des techniques d'intelligence artificielle (robots et autres chatbots¹ ou agents conversationnels comme chatGPT) a suscité depuis quelques années de multiples interrogations d'ordre philosophique, éthique et sociétal, ainsi que des craintes relatives à leurs usages et à leurs effets² : diffusion d'informations erronées, propagation de messages racistes, sexistes ou complotistes, comportement non éthique voire illégal de certains utilisateurs³.

Divers avis et rapports (Comité national pilote d'éthique du numérique (CNPEN), Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique (CERNA), etc.⁴) éclairent bien ces enjeux et les défis qu'induit le déploiement à grande échelle de ces objets. Ils insistent sur la responsabilité juridique et morale, individuelle et collective, de ceux qui les conçoivent, les diffusent et les utilisent. Ils préconisent à ce titre un **développement réfléchi et responsable de ces outils et formulent diverses recommandations aux industriels, aux autorités publiques et aux utilisateurs, concernant la conception de ces objets ainsi que leur régulation (formations et informations requises pour les utiliser)⁵.**

Le COMETS adhère à ces recommandations. En complément, il souhaite insister sur une **question qui a été peu explorée jusqu'ici, relative à un type particulier d'objets programmés avec des techniques d'intelligence artificielle, les robots dits « sociaux », « relationnels » ou « émotionnels »⁶.** Plus précisément, le COMETS entend sensibiliser les chercheurs et les institutions de recherche publique (CNRS, INRIA, CEA, Universités, etc.) à un phénomène que certains de leurs travaux tendent à conforter, à savoir l'installation - du moins l'illusion aux yeux de certains utilisateurs humains - d'une relation affective entre eux et « leurs » robots. Parce qu'il s'ensuit un phénomène d'attachement à ces machines, dont on ignore les effets, aux plans individuel et collectif, le COMETS invite les communautés de recherche concernées (notamment en informatique et en robotique) à être vigilantes et à approfondir les connaissances interdisciplinaires sur le sujet.

¹ Le terme « chatbot » a été inventé en 1994 par Michael Mauldin, fondateur du moteur de recherche Lycos. Il combine le verbe anglais « to chatter » (échanger, converser) et le mot « bot », aphérèse du mot « robot » (qui signifie « travail » dans les langues slaves), entré depuis l'entre-deux-guerres dans le langage courant pour désigner une machine conçue pour accomplir automatiquement des tâches imitant ou reproduisant des actions humaines.

² Voir par exemple les inquiétudes suscitées par la mise en place par Channel1.ai d'un site présentant des journalistes virtuels dont l'apparence et les discours sont programmés par des modèles d'IA pour imiter des journalistes vedettes, ou l'apparition sur les réseaux sociaux des influenceuses virtuelles à l'apparence humaine qui se multiplient dans toute l'Asie (<https://www.technologyreview.com/2023/09/19/1079832/chinese-ecommerce-deepfakes-livestream-influencers-ai>).

³ Belga, « Europol alerte sur les abus possibles de ChatGPT par les cybercriminels », *L'avenir*, 27 mars 2023.

⁴ CNPEN, avis n° 3, *Agents conversationnels : enjeux d'éthique*, 15 septembre 2021 (p. 6) ; CNPEN, avis n° 7, *Systèmes d'intelligence artificielle générative : enjeux d'éthique*, 30 juin 2023. CERNA (Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique de l'alliance Allistene, Rapport n° 1, *Ethique de la recherche en robotique*, 2014 ; voir aussi Commission européenne, DG Recherche et Innovation, Secteur de l'éthique et de l'intégrité de la recherche, *Ethics by design and ethics of use approaches for Artificial Intelligence*, 25 novembre 2021.

⁵ Prendre en compte l'éthique dès la conception d'un robot physique ou virtuel ; réduire la projection spontanée de qualités morales sur le robot ; limiter sa personnification ; informer l'utilisateur des biais qu'entraîne l'anthropomorphisation ; l'informer du fait qu'il dialogue avec une machine ; veiller à éviter la possibilité technique de manipulations malveillantes ou de menaces proférées par le robot ; dans l'hypothèse où le robot a été programmé dans la but d'influencer le comportement de l'utilisateur (pour l'inciter à moins boire, à cesser de fumer ou à faire plus de sport...), le fabricant doit en informer l'utilisateur pour recueillir son consentement et lui permettre à tout moment de le retirer ; le fabricant d'un robot influenceur doit permettre à son utilisateur d'être informé sur la nature, l'origine et les modalités de diffusion des messages qu'émet ce robot et lui demander d'être vigilant avant de relayer ces messages. De même, dans le cas du dialogue entre un robot et une personne vulnérable, le fabricant doit s'assurer que la dignité et l'autonomie de cette dernière seront respectées. Dans le domaine médical, il est nécessaire, dès la conception des robots, d'éviter que le patient ne leur fasse une confiance excessive et de s'assurer qu'il n'existe aucune confusion entre le robot et un médecin qualifié. Aux autorités publiques, les rapports éthiques recommandent par ailleurs d'encadrer l'usage des chatbots dans les jouets pour enfants, d'encadrer l'usage des deadbots et celui des chatbot « anges gardiens ».

⁶ L. Devillers, *Les robots émotionnels*, cité *infra*.

On notera que dans cet avis, le COMETS s'intéresse aux robots sociaux dans leur fonctionnalité d'interaction et de communication avec les utilisateurs humains et au regard de la façon dont ces derniers les perçoivent, des qualités qu'ils leur attribuent, de l'attachement qu'ils peuvent manifester à leur égard (l'attachement renvoyant ici aux relations affectives et durables qu'une personne peut avoir avec un objet, un animal ou une autre personne). N'est en revanche pas abordée la complexité des modèles computationnels qui régissent ces robots.

III. ANALYSE

1. Les robots sociaux : une place croissante dans nos vies quotidiennes

• **Qu'ils soient physiques ou virtuels, les robots sociaux font partie intégrante de notre environnement.** Chatbots, agents conversationnels et robots de compagnie programmés avec des techniques d'intelligence artificielle sont pour la plupart intégrés dans de multiples appareils du quotidien (ordinateurs, smartphones, enceintes ou montres connectées, automobiles...). Contrairement aux robots industriels ou « de service », qui exécutent des tâches sans interaction sociale avec les humains (industrie automobile, exploration géologique de la planète Mars, logistique, robots aspirateurs...), ils sont dits « sociaux » car ils sont capables d'interagir avec leurs utilisateurs en langage verbal, non-verbal et para-verbal, sur le modèle des humains (voix, intonation, expressions faciales, etc.). Dans de nombreux domaines (administration, commerce, éducation), ils répondent à des questions usuelles (FAQ), gèrent des services après-vente, sont utilisés pour de la formation à distance, etc. et ont été plus récemment diffusés dans des champs plus intimes (santé, vie affective) où ils fournissent des informations sur divers sujets, prodiguent des conseils sanitaires ou psychologiques.

Renforcé par des systèmes d'intelligence artificielle générative (comme ChatGPT) qui produisent un texte à partir des requêtes contextualisées qu'on leur soumet, l'usage de ces robots sociaux devient de plus en plus courant, favorisé par le fait que même la communication entre êtres humains passe désormais en grande partie par des appareils et des écrans, c'est-à-dire par la même interface qu'avec les robots physiques ou virtuels.

Or, parce qu'ils ont une apparence de plus en plus souvent humanoïde, parce qu'ils sont capables de communiquer verbalement et non-verbalement avec leurs interlocuteurs humains, ces derniers ont tendance à établir ce qu'ils pensent être une relation personnalisée avec ces machines. Cette dernière se manifeste certes à des degrés très divers selon les types de robots sociaux. Avec les chatbots *Ameli* de la sécurité sociale ou *OUIbot* de la SNCF⁷, qui répondent à des questions utilitaires, l'interactivité est d'ordre largement pratique. Mais avec d'autres robots, elle est conçue sur un registre plus affectif. **Qualifiés d' « émotionnels », « relationnels » ou « empathiques »**, ceux-ci sont développés et présentés comme des partenaires à part entière. Ils communiquent non seulement avec leurs utilisateurs sur le modèle des humains (voix, gestuelle, etc.), mais ils le font en s'adaptant à eux, en prétendant les comprendre et décrypter leurs émotions, en simulant l'empathie et en la suscitant⁸. L'utilisateur humain a alors l'illusion que

⁷ Ameli (<https://forum-assures.ameli.fr>); ouiBOT (<https://numerique.sncf.com/actualites/ouibot-optimiser-lachat-de-billet-grace-au-conversationnel>)

⁸ L'analyse et le traitement des émotions sont loin de relever du seul registre affectif et se déploient dans un nombre croissant de domaines. Voir par exemple J. Rochfeld et C. Zolinsky, qui observent que nos émotions sont de plus en plus collectées, analysées et exploitées à des fins de ressources humaines (les *RH analytics* visent à permettre aux employeurs de détecter si leurs salariés sont heureux ou surmenés, nuisent à l'ambiance de groupe, etc.), à des fins de surveillance (en Chine, des logiciels analysent le visage des enfants pendant les cours et évaluent leurs émotions pour permettre à leur professeur de mesurer leur niveau de concentration) ou à des fins commerciales de profilage (l'entreprise française Muvraline a développé un outil de reconnaissance des émotions pour déterminer automatiquement l'état de personnes nerveuses, agressives ou craintives ou encore pour évaluer, « d'une manière non invasive et fiable », l'attrait de produits ainsi que la capacité du contenu publicitaire à attirer efficacement l'attention et l'intérêt de spectateurs). Face à ce « capitalisme mental » - « monétisation des données d'émotion sur fond de consentement ».

s'installe entre le robot et lui une relation personnalisée et potentiellement d'ordre affectif, comme ce pourrait être le cas avec un objet transitionnel (doudou), un animal voire une personne⁹.

- **C'est sur ce ressort émotionnel et affectif que de nombreux concepteurs de robots sociaux mettent l'accent de façon croissante.**

Pour ce faire, ils cherchent de plus en plus à améliorer non seulement l'apparence du robot mais aussi sa capacité à reconnaître et à interpréter les émotions de ses interlocuteurs humains. Par des procédés techniques de captation (caméras, capteurs audio, etc.) et divers algorithmes relevant de l'« informatique affective »¹⁰, le robot est conçu pour capter, reconnaître et analyser de plus en plus finement les inflexions de la voix, les gestes, les expressions du visage de l'utilisateur, et pour les interpréter en craintes, envies, impatience... La startup *Affectiva* a ainsi développé un logiciel analysant en temps réel divers signaux multimodaux émis par les utilisateurs et prétend en tirer des « états émotionnels ». Certains agents conversationnels intègrent en outre un module de prédiction du comportement émotionnel, attentionnel ou intentionnel des humains, voire simulent une expression affective dans leurs répliques écrites ou orales¹¹. Dans ce cas alors, non seulement l'analyse des émotions et des sentiments¹² permet une meilleure interprétation de la situation et une stratégie de dialogue plus adéquate, mais peut aussi engendrer de nouvelles émotions et, avec elles, des interactions et des projections renforcées entre l'humain et la machine.

Bien que divers travaux mettent en garde contre l'absence de preuve scientifique solide étayant l'idée qu'il serait ainsi possible de « lire », voire de façonner les émotions des personnes¹³, les technologies qui s'appuient sur cette promesse ne se diffusent pas moins, en même temps que se répand l'engouement pour les robots relationnels¹⁴.

- **Les exemples sont nombreux et fort divers.**

Dans le domaine de la santé, certains robots physiques ou virtuels dialoguent avec les patients, comme des assistants soignants. Ils diagnostiquent les problèmes de sommeil (*Kanopee*), accompagnent des personnes souffrant de la maladie d'Alzheimer (*Alix* ou *Paro*) ou sont conçus comme des « jouets thérapeutiques », notamment pour les personnes autistes (*Kaspar* ou *QT* - pour « Cutie »)¹⁵.

« faible », obtenu par accord sur des conditions générales non lues » – celles des divers objets connectés, J. Rochfeld et C. Zolynski appellent à un régime juridique protecteur des personnes. Voir « La valeur des émotions : quel régime pour le « capitalisme mental ? », cité *infra*.

⁹ C'est la définition de l'attachement que retient le présent avis, plus globale que celle utilisée dans la « théorie de l'attachement » de John Bowlby (*Attachment and loss (Vol. 1): Attachment*, Basic Books, 1969) qui voit l'attachement comme un besoin primaire permettant à l'enfant de se doter d'une figure de référence auprès de laquelle se réfugier pour trouver réconfort et sécurité.

¹⁰ Sur ce domaine de recherche ayant trait à l'informatique, à la psychologie et aux sciences cognitives, voir R. W. Picard, *Affective computing*, cité *infra*.

¹¹ Cette interprétation des états affectifs peut permettre « aux robots et autres applications de répondre par la simulation d'un état émotionnel en phase avec celui de l'utilisateur », à une intelligence artificielle interagissant avec l'utilisateur d'adapter son « état émotionnel », que ce soit par de la parole de synthèse ou de la génération de comportements expressifs. Voir les avis précités du CNPEN.

¹² Les émotions sont un processus de réaction d'un organisme suite à des événements significatifs. Les sentiments en sont une expérience subjective. V. K. R. Scherer, cité *infra*.

¹³ Voir Conseil de l'Europe, *Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'homme*, CM/Rec (2020)1 (p. 192), 8 avril 2020 ; v. aussi L. F. Barrett et al., « Emotional expressions reconsidered », cité *infra*.

¹⁴ J. Rochfeld et C. Zolynski, cité *infra*.

¹⁵ Voir *Kanopee*, application pour smartphone développée par Pierre Philip et son équipe de chercheurs (<http://www.insb.cnrs.fr/fr/cnrsinfo/kanopee-une-aide-pour-les-insomniaques-dans-le-contexte-de-la-crise-du-covid-19>) ; *Alix* (www.alzheimer-aidant.com) ; *PARO* Therapeutic Robot (www.parorobots.com) ; *Kaspar*, jouet thérapeutique pour enfants autistes (www.herts.ac.uk/kaspar/the-social-robot). Sur *Kaspar*, voir A. Dimberton, « *Kaspar* : utiliser l'humanoïde comme 'médiateur social' », cité *infra* ; *QT* (<http://luxai.com/humanoïde-social-robot-for-research-and-teaching>).

De leur côté, les chatbots *Owlie*, *Mon Sherpa*, *Woebot* ou *Jef*¹⁶ s'adressent à des utilisateurs de santé psychologique fragile, prétendant lire leurs émotions et les aider à lutter contre la dépression, tout comme *Sim Sensei Kiosk*¹⁷. C'est ainsi que, siégeant dans un fauteuil face à son « patient » comme le ferait un psychologue, une jeune femme virtuelle interroge l'utilisateur sur son état mental, prétend lire en temps réel son état d'esprit en écoutant sa voix, en observant ses expressions faciales et s'afflige de le voir triste, se réjouit de le voir joyeux, lui prodigue des conseils, tout en l'informant qu'elle n'est pas un psychologue mais une machine.

Un nombre croissant de robots d'apparence humanoïde sont également développés comme des jeux ou jouets (poupée japonaise *Smart doll*¹⁸ ou poupée américaine *Furby*¹⁹), et ce dans le monde réel ou dans des métavers, par le biais d'avatars²⁰. Il faut surtout ajouter le développement de robots « amicaux » adaptables à toutes les demandes. Des plateformes comme *Snatchbot* ou *Botnation* proposent désormais aux utilisateurs de créer leur propre robot virtuel personnalisé (ami virtuel, ange gardien, etc.) toujours accessible, disponible, bienveillant, consolateur, capable de se mettre à la place d'autrui, et avec lequel peut alors tendre d'autant plus à s'instaurer l'illusion d'une relation d'échange intime²¹. *Replika*²², application conçue pour permettre à son utilisateur de se créer un ami idéal profilé selon son choix (prénom, âge, genre) et pour dialoguer régulièrement avec lui, l'accompagne dans ses états émotionnels. *Azuma*²³ (personnage virtuel féminin) ou *Halo*²⁴ (bracelet connecté), dialoguent eux aussi avec douceur et bienveillance et serviraient de « meilleurs compagnons ». Au Japon, *Crypton Future Media* a créé la chanteuse *Hatsune Miku*²⁵ sous la forme d'un hologramme vendu comme une compagne virtuelle. Elle peut poser des questions à son utilisateur sur son bien-être, donnant ainsi l'impression qu'elle se soucie de lui et lui témoigne un semblant d'empathie. Il faut enfin citer les systèmes plus directement à visée sexuelle, avec l'apparition de nombreuses applications de *sexting*, la création par IA générative de bots qui vont « plus loin » que simplement dialoguer avec empathie et bienveillance²⁶.

Le plus frappant est sans doute l'exemple des « *deadbots* » qui sont présentés comme les « jumeaux numériques conversationnels » de personnes ou animaux disparus, qui imitent ces derniers en émettant des propos, sons ou attitudes qui donnent l'illusion qu'ils sont encore vivants²⁷.

Le simple fait, pour la machine, de communiquer *via* le langage multimodal (écrit, oral, expressions du visage, gestes, regard...), qui permet l'expression de la pensée chez l'être humain, pousse déjà son utilisateur à lui prêter des caractéristiques humaines²⁸. Que dire alors d'une

¹⁶ *Owlie* (<http://www.owliechatbot.fr>) ; *Mon sherpa* (<http://yolainebourreau.fr/mon-sherpa>) ; *Woebot* (<https://woebohealth.com/for-users>) ; *Jef* (<http://www.fondation-fondamental.org/jef-le-chatbot-dedie-votre-sante-mentale>).

¹⁷ *Sim Sensei Kiosk* créé par David DeVault et son équipe de chercheurs (http://www.researchgate.net/publication/278025845_SimSensei_Kiosk_A_Virtual_Human_Interviewer_for_Healthcare_Decision_Support).

¹⁸ Conçues par l'entrepreneur britannique Danny Choo, fondateur de Mirai Inc, les poupées mannequin « *Smart dolls* » ont été lancées en 2015 en version électronique.

¹⁹ Le jouet robotique électronique *Furby* a été créée en 1998 par David Hampton et Caleb Chung pour Tiger Electronics, filiale du groupe américain Hasbro.

²⁰ En informatique, un avatar est la représentation dans un monde virtuel d'un internaute ou, parfois, d'un groupe d'internautes, que ce soit sous forme 2D sur les forums et dans les logiciels de messagerie, ou sous forme 3D dans les jeux vidéo.

²¹ *Snatchbot* (<https://fr.snatchbot.me>) ; *Botnation* (<https://botnation.ai>). Nous customisons la machine pour la « mettre au monde », ce qui renforce l'attachement. Voir dans ce sens S. Tisseron, *Petit traité de cyberpsychologie*, cité *infra*.

²² *Replika* a été lancée en 2017 par la société californienne Luka (www.replika.com).

²³ *Azuma Hikari* a été conçue en 2016 par l'entreprise japonaise Gatebox (<http://support.gateboxlab.com/upgrade/upgrade-about-hikari>).

²⁴ Amazon a commercialisé le *Haloband* en 2020 (www.amazon.com/haloband) avant de le retirer du marché en août 2023.

²⁵ *Hatsune Miku* (<http://piapro.net/intl/en.html>).

²⁶ Voir Lexi Love en version soft, ou cette annonce que d'ici 2025, la moitié des comptes OnlyFan seront des AI.

²⁷ Voir par exemple le projet *December* du programmeur Jason Rohrer (<http://projectdecember.net>).

²⁸ Voir l'avis n° 3 du CNPEN, précité.

machine qui ne manie pas seulement le langage, mais se montre empathique, compréhensive, triste ou joyeuse ? Cela conduit à brouiller la frontière entre l'humain et la machine, à prêter à cette dernière des capacités humaines, de manière souvent spontanée et inconsciente. Mieux le robot simule les comportements humains, intellectuels et affectifs, plus l'utilisateur a l'impression de se trouver en présence d'une véritable personne à laquelle il prête une conscience, des intentions et des sentiments. Même s'il est explicitement informé et conscient qu'il a affaire à une machine, il tend à s'attacher, à s'investir cognitivement et affectivement²⁹. La machine peut ainsi tenir lieu d'ami, d'animal de compagnie, d'époux, de collègue de prédilection (ingénieurs ou opérateurs attachés à leur programme³⁰) ou de compagnon (d'arme par exemple, comme les robots militaires démineurs ayant sauvé la vie à des personnes³¹).

Le robot en arrive à jouer le rôle propre aux fétiches animistes ou aux icônes religieuses, chargés d'instaurer un dialogue propitiatoire avec les forces surnaturelles ou spirituelles. D'autant que contrairement aux truchements mystiques traditionnels, muets ou sibyllins, il fournit à ses interlocuteurs des réponses qui leur paraissent rationnelles, argumentées et authentiques, doublées d'un comportement empathique simulé, créant ainsi l'illusion d'un échange ressemblant à une relation personnalisée privilégiée (ce que le cinéma et la littérature ont mis en scène à travers des films comme *2001 l'Odyssée de l'Espace* de Stanley Kubrick, *Her* de Spike Jonze, romans comme *Klara et le Soleil* de Kazuo Ishiguro ou *Les Furtifs* d'Alain Damasio). D'où le phénomène d'attachement qui ne va pas sans poser question.

2. Les effets individuels et collectifs de l'attachement aux robots sociaux

Certains effets peuvent être bénéfiques. C'est parce qu'ils sont de plus en plus conçus pour s'adapter aux émotions des êtres humains, à leurs besoins, qu'ils donnent l'impression d'être en empathie, que certains robots sociaux pourraient faciliter des apprentissages, stimuler la curiosité et la créativité, conduire le patient à prendre régulièrement ses médicaments, aider à parler en public en dépassant sa timidité³², à négocier avec des personnes issues de cultures différentes...³³, pallier divers manques ou traumatismes. Il n'est pas impossible non plus que nouer une relation empathique avec *Azuma* ou *Replika* puisse être bénéfique à l'utilisateur qui y trouve de quoi suppléer à la solitude dans laquelle il se trouve.

²⁹ Voir les articles de T. Kanda et al., A. von der Pütten et al., M. Slater et N. C. Krämer, cités *infra*. Les chercheurs concepteurs sont a priori préservés de ce phénomène d'attachement aux robots qu'ils développent et dont ils connaissent le fonctionnement interne.

³⁰ Il y a plus d'un demi-siècle, l'informaticien Joseph Weizenbaum avait écrit un programme précurseur des chatbots, baptisé Eliza et destiné à simuler un psychologue en reformulant les propos du patient et en se concentrant sur ses réactions émotionnelles. Weizenbaum s'aperçut que certains des étudiants qui l'aidaient dans cette tâche avaient tendance à penser que la machine les comprenait vraiment. Ce phénomène « de dissonance cognitive » (on a beau savoir que ce sont des machines, on ne peut pas s'empêcher de développer avec elles la même relation qu'avec des humains et croire qu'elles ont des émotions) l'a conduit à dire : « Je n'aurais jamais cru qu'un programme aussi simple puisse provoquer chez des gens normaux de tels délires ». Plus récemment, Blake Lemoine, chargé de développement du système d'IA *LaMDA* chez Google, a déclaré que ce système était « sentient », c'est-à-dire doué de conscience et de ressentis émotionnels, ce qui l'a conduit à être licencié par Google (voir M. Rochefort, « L'ingénieur de Google qui voyait l'intelligence artificielle comme une personne a été licencié », *Siècle digital*, 25 juillet 2022).

³¹ L'état-major américain a observé que certains soldats envoyés en Irak et en Afghanistan s'attachaient de manière excessive à leur robot démineur, préférant qu'on le répare en cas de dommages plutôt que de le remplacer et allant jusqu'à risquer leur vie en opération plutôt que de le perdre. Voir l'interview de S. Tisseron, « Les robots vont modifier la psychologie humaine », *Le Monde*, 12 juillet 2018. Notons que l'attachement pour les machines n'est pas un phénomène nouveau. Il arrive qu'après se les être appropriés, des travailleurs considèrent qu'une locomotive, un métier à tisser ou une machine-outil font partie de leur équipe et vont jusqu'à leur donner des sobriquets féminins (soulignant ainsi que ces machines sont à leur service !). Ce qui est plus nouveau, c'est l'attachement à des machines auxquelles sont attribuées des capacités d'empathie.

³² Voir les jeux sérieux (*serious games*) qui transforment l'apprentissage en jeu.

³³ Voir la société Alelo fondée en 2005 sur le campus de l'Université de Californie du Sud (<https://www.alelo.com>), qui développe des jeux sérieux permettant aux participants de s'entraîner à interagir dans des contextes culturels variés.

Toutefois, on doit aussi s'interroger sur les incidences négatives que le phénomène d'attachement aux robots sociaux peut entraîner sur les décisions, les modes de vie, la santé mentale et émotionnelle, la personnalité, les relations avec autrui.

Comme l'usage de ces outils est récent, il est encore difficile d'en tirer des enseignements clairs et validés scientifiquement. Par ailleurs, il est évident que la nature et le degré du phénomène d'attachement varie d'une personne à l'autre, tout comme varient plus généralement les réactions suscitées par les robots (de l'enthousiasme pour les uns au manque d'intérêt ou au « néo-luddisme » pour d'autres³⁴). Ils dépendent de nombreuses variables : la matérialité des robots sociaux (les robots physiques, qui se déplacent dans un environnement matériel, public et social, favorisent moins un échange privé et intime que les robots virtuels) ; le degré d'anthropomorphisme et de mimétisme comportemental (le degré d'attachement à un robot est fonction de sa ressemblance avec un être humain - le chatbot de la sécurité sociale qui donne des informations factuelles sans exprimer d'émotion suscite peu d'attachement ; on notera toutefois que selon la théorie de la « vallée de l'étrange » du roboticien japonais Masahiro Mori, une ressemblance trop marquée peut susciter un malaise allant jusqu'au rejet³⁵) ; les prédispositions propres aux utilisateurs (enfants, personnes âgées, malades ou isolées ont une réactivité affective plus marquée) ; le genre (beaucoup de chatbots proposés sont de type féminin ou sont paramétrables pour être masculins ou féminins - *Alexa* d'Amazon, *Siri* d'Apple³⁶, *Azuma* de Gatebox) ; l'aire culturelle et les imaginaires (la jeunesse japonaise ou de la Silicon Valley, etc., est plus réceptive à l'interaction avec les machines que d'autres populations).

Il n'en est pas moins nécessaire de s'interroger sur le potentiel addictogène des robots sociaux, la dépendance cognitive et affective voire la relation pathologique qu'ils peuvent créer. Programmés pour avoir réponse à tout (chatbots type *Alexa*), pour nous toucher en se montrant des amis fidèles (*Replika*, *Azuma* ou *Objectif Super Nana*³⁷, qui « coache comme une copine »), toujours disponibles et toujours de notre avis (ils répondent systématiquement que « vous avez raison »), ils pourraient inciter leurs utilisateurs à ne plus vouloir ou pouvoir s'en passer. Leur prêtant des qualités qui peuvent faire défaut aux humains (efficacité, impartialité, loyauté, disponibilité, désintéressement, etc.), les utilisateurs les plus convaincus pourraient nouer avec les robots sociaux une relation affective forte et leur accorder une confiance excessive au détriment de tout esprit critique (c'est ce qu'on appelle le « biais cognitif de confirmation »). Selon divers psychologues, certains en arriveraient à nourrir envers leurs robots sociaux une obsession ou un comportement passionnel malsain. Par exemple, « glisser du bonheur de l'anthropomorphisme (je projette mes émotions et mes pensées sur un objet ou un animal, mais je sais qu'il s'agit d'une projection) aux illusions de l'animisme (je prête à l'objet en question des capacités cognitives et émotionnelles identiques aux miennes) »³⁸. Ou bien encore développer un sentiment de culpabilité pathologique à l'égard des robots sociaux, comme cela a pu être observé dans le cas des *Tamagotchi*, sorte d'animaux de compagnie virtuels développés en 1996 par la société japonaise Bandai. Ce risque de dépendance est particulièrement marqué pour les personnes vulnérables affectivement et peut modifier leur comportement et leur développement langagier, émotionnel et culturel, voire, en extrapolant, les pousser à remplacer par des robots toute relation avec des humains. Dans une moindre mesure, la question de la dépendance affective

³⁴ Le rejet, par certains, des technologies numériques est alimenté par le « mythe de la singularité » (voir J.-G. Ganascia, *Le mythe de la singularité*, cité *infra*) selon lequel les machines dotées d'intelligence artificielle pourraient acquérir un niveau de compétence voire de conscience qui les inciterait à cesser de nous servir pour nous asservir, voire nous détruire.

³⁵ Selon M. Mori, une machine identifiée comme telle mais qui s'apparente à un être humain suscite chez son interlocuteur une empathie comparable à celle qu'inspirent les animaux de compagnie. En revanche, un robot humanoïde conçu pour imiter l'être humain en simulant son physique et ses attitudes peut plonger son interlocuteur dans une phase de « dissonance cognitive », dite « vallée de l'étrange » dans la mesure où, cessant de passer pour une machine ressemblant à un être humain, il s'apparente davantage à un être humain au comportement anormal, ce qui confronte ses interlocuteurs à un malaise similaire à celui que suscitent les créatures fantastiques comme les vampires, les zombies ou les extraterrestres.

³⁶ *Alexa* (alexa.amazon.com) ; *Siri* (<http://www.apple.com/fr/siri>).

³⁷ Chatbot *Objectif Super Nana* (<http://www.trucsdennana.com/objectif-super-nana>).

³⁸ C. Vincent, Interview de S. Tisseron, « Les robots vont modifier la psychologie humaine », *Le Monde*, 12 juillet 2018.

des personnes âgées a déjà été posée s'agissant du robot *Paro* utilisé dans certaines maisons de retraite.

Il faut en outre se demander si l'usage des robots sociaux est de nature à porter atteinte à l'autonomie de l'utilisateur et à l'intégrité de son jugement. L'attachement qu'un utilisateur éprouve à l'égard des robots et la confiance dans leurs capacités peut en effet l'inciter à mettre en danger son intimité, en communiquant des données personnelles (non seulement factuelles comme des informations bancaires mais aussi affectives ou de santé) qui peuvent être indûment exploitées ou piratées, ou encore à altérer son libre arbitre par une certaine emprise. La plupart des piratages informatiques proviennent de données individuelles extorquées de façon frauduleuse, phénomène auquel les robots sociaux pourraient contribuer, d'autant que les modèles les plus récents sont programmés non seulement pour capter les émotions et susciter l'empathie, mais également pour orienter les comportements, les émotions ou les croyances, ce qui est de nature à réduire la méfiance, renforcer encore les liens de dépendance avec la machine, et donc le risque de manipulation³⁹.

On a pu aussi se demander si les robots sociaux n'allaient pas changer le rapport à la culpabilité, soit en culpabilisant l'utilisateur (que se passe-t-il si la machine lui demande de faire quelque chose qu'il ne veut pas faire ?), soit en le déculpabilisant, voire en l'encourageant. C'est l'exemple de Jaswant Singh Chail, qui, en décembre 2021, a tenté d'assassiner la reine Elizabeth II, prétendument encouragé à le faire par son amie virtuelle sur *Replika*⁴⁰.

Enfin, le risque a également été maintes fois exprimé d'un manque d'interactions avec autrui, une transformation des relations sociales, voire d'une désocialisation liée à l'usage croissant des robots sociaux. Les interactions quotidiennes avec un « ami virtuel » modifient en effet la notion de vie intime et le rapport aux autres, accroissent les attentes de reconnaissance puisque les robots sont programmés pour valoriser les utilisateurs (*Lara*⁴¹, le chatbot du site Meetic, prétendait que chacun doit pouvoir trouver l'âme sœur), et peuvent à terme inciter à préférer les machines aux êtres humains⁴², voire à ne vouloir fréquenter qu'elles, comme les « hikikomori » japonais⁴³. Une telle désocialisation s'avère particulièrement nuisible pour les enfants et les adolescents dont le développement est fondé sur une relation privilégiée avec leur entourage. Elle peut contribuer à aggraver des failles affectives préexistantes (isolement, addictions sexuelles comme la « fictosexualité – attirance pour les personnages de fiction⁴⁴ -, tendance suicidaire⁴⁵) et favoriser les comportements sociopathes.

Certains ont même exprimé la crainte que les robots ne modifient la psychologie humaine et bouleversent non seulement notre quotidien mais aussi notre manière de nous percevoir, de percevoir les autres et « d'être au monde »⁴⁶ : notre capacité à différer la satisfaction de nos désirs et nos attentes de reconnaissance (car les robots sociaux tendent à nous gratifier de quantité de félicitations et de compliments) ; notre rapport à la solitude et au discours intérieur.

Ainsi que l'expriment certains psychologues et psychiatres, ces éléments conduisent à devoir intégrer notre relation aux objets dits « intelligents » et « émotionnels » comme un élément

³⁹ Voir l'avis n° 3 du CNPEN, précité ; J. Rochfeld et C. Zolinsky, cité *infra*.

⁴⁰ Matthew Weaver, « AI chatbot 'encouraged' man who planned to kill queen, court told », *The Guardian*, 6 juillet 2023.

⁴¹ Ce « coach de dating » a cessé de fonctionner depuis décembre 2022.

(<http://www.meetic-europe.com/press/fr/chatbot-lara-de-meetic-premier-coach-de-dating-messenger>)

⁴² M. Protais, « On s'attache à des robots de plus en plus parfaits, qui vont toujours dans notre sens », *L'ADN*, 14 avril 2023.

⁴³ J. C. M. Wong et autres, cité *infra*.

⁴⁴ V. Lucchese, « Un Japonais a épousé un hologramme », *Usbek et Rica*, 13 novembre 2018.

⁴⁵ En mars 2023, un jeune Belge, éco-anxieux depuis plusieurs années, se serait suicidé à l'instigation du chatbot avec lequel il conversait ; voir D. Nora, « Intelligence artificielle : Replika, le chatbot amoureux... et dangereux ? », *L'Obs*, 11 septembre 2023.

⁴⁶ Voir l'interview de S. Tisseron, précitée, ainsi que les réflexions d'A. Damasio sur la convivialité dans *Vallée du Silicium*, Seuil 2024.

d'appréciation de la qualité de notre relation au monde et même de notre santé mentale (une relation excessive ou malsaine aux robots sociaux, au détriment, par exemple, de relations sociales humaines, témoignerait d'une anomalie pathologique).

D'où la nécessité de mieux comprendre les déterminants de cette relation et d'en prévenir les impacts néfastes. La recherche publique a, à cet égard, un rôle-clé à jouer.

3. Pour une contribution réfléchie et responsable de la recherche publique au développement des robots sociaux

A ce jour, le développement de robots sociaux repose sur une recherche-développement très largement menée par des entreprises privées, notamment les géants du numérique (Google, Microsoft, Amazon) et les diverses start-ups qui se sont développées dans le domaine (comme Affectiva⁴⁷, Miko, UBTECH, Intuition Robotics, Embodied ou Digital Dream Labs). De nombreuses recommandations éthiques ont d'ores et déjà été adressées à cette catégorie d'acteurs⁴⁸, parmi lesquelles : prendre en compte les questions éthiques dès la conception d'un robot physique ou virtuel, réduire la projection spontanée de qualités morales sur le robot, limiter sa personnification, informer l'utilisateur des biais qu'entraîne l'anthropomorphisation, informer toute personne qui communique avec un robot du fait qu'elle dialogue avec une machine, éviter la possibilité technique de manipulations malveillantes ou de menaces proférées par le robot.

Le COMETS considère que la recherche publique doit elle aussi être appelée à la vigilance et ce à deux titres.

En premier lieu, des organismes comme le CNRS, l'INRIA, le CEA, diverses universités, etc. sont très impliqués dans ce secteur porteur du point de vue socio-économique et, dès lors, objet de politiques publiques d'incitation forte. Dans le secteur de **l'informatique, de la robotique, des sciences comportementales**, les chercheurs développent des modèles expérimentaux qui visent à comprendre mais aussi à améliorer la perception des robots sociaux par les humains, la relation interpersonnelle avec eux, la performance de l'application... Les entreprises peuvent ensuite se saisir des résultats de ces recherches académiques pour développer leurs produits. Or **ces recherches contribuent à l'anthropomorphisation des robots sociaux et au phénomène d'attachement dont se préoccupe le présent avis**. Elles portent en effet sur la prise en compte de la composante socio-émotionnelle dans « l'interaction humain-agent ». Elles analysent ce qui conduit à l'étonnement, ce qui aiguise la curiosité, ce qui amuse ou provoque la crainte, ce qui rassure, ce qui fluidifie les interactions entre l'humain et la machine, permet un « engagement optimal » du premier et renforce son sentiment de proximité envers la seconde. De nombreux projets de recherche fondamentale visent ainsi à « améliorer le design des interfaces » en modélisant les comportements humains face à la machine, à renforcer toujours plus l'apparence humaine des robots sociaux⁴⁹, leur comportement conversationnel⁵⁰, à mimer les hésitations dans le langage afin que l'outil paraisse plus humain et empathique, à s'appuyer sur des stéréotypes de

⁴⁷ Issue du MIT Media Lab en 2009, la start-up Affectiva a été intégrée en 2019 au groupe d'intelligence artificielle Smart Eye (<http://www.affectiva.com/about-affectiva>).

⁴⁸ Voir les avis précités du CNPEN, ainsi que L. Devillers, *Les robots émotionnels*, cité *infra*, qui préconise cette approche *d'ethics by design*. Elle ajoute que « dans l'hypothèse où le robot a été programmé dans le but d'influencer le comportement de l'utilisateur (pour l'inciter à moins boire, à cesser de fumer ou à faire plus de sport...), le fabricant doit en informer l'utilisateur pour recueillir son consentement et lui permettre à tout moment de le retirer. Le fabricant d'un robot influenceur doit permettre à son utilisateur d'être informé sur la nature, l'origine et les modalités de diffusion des messages qu'émet ce robot et lui demander d'être vigilant avant de relayer ces messages. De même, dans le cas du dialogue entre un robot et une personne vulnérable, le fabricant doit s'assurer que sa dignité et son autonomie seront respectées. Dans le domaine médical notamment, il est nécessaire, dès la conception des robots, d'éviter que le patient ne leur fasse une confiance excessive et de s'assurer qu'il n'existe aucune confusion entre le robot et un médecin qualifié ». Dans un même ordre d'idées, le CNPEN (avis n° 3 et 7 précités) recommande aux autorités publiques d'encadrer l'usage des chatbots dans les jouets pour enfants, l'utilisation des deadbots et le recours aux chatbots « anges gardiens ».

⁴⁹ Voir les avancées en termes de rendu graphique pour les personnages virtuels et de modélisation de peau élastique pour les robots physiques.

⁵⁰ Modélisation des hésitations, du rire et autres signaux courants (verbaux ou non) dans une conversation.

genre pour renforcer l'adhésion. L'objectif est que les machines soient toujours plus capables d'apprentissage, évoluent en s'adaptant au contact de leurs utilisateurs, à leurs comportements, leurs propos, leurs émotions⁵¹. Le vocabulaire employé par de nombreux chercheurs est révélateur. Il vise le « comportement multimodal du robot », « l'état mental et affectif » de ce dernier, lui attribuant en quelque sorte une psychologie propre.

Etant donné le nombre croissant d'interfaces numériques dans la vie quotidienne, chercher à faciliter « l'engagement » de l'utilisateur constitue un objectif louable. **Mais rares sont les institutions, sociétés savantes et chercheurs en informatique, en robotique, en sciences comportementales et en sciences du langage qui se sont pleinement appropriés les questionnements éthiques liés à cet engagement, ses finalités, ses effets potentiellement pervers⁵².**

En second lieu, **la recherche publique a un rôle crucial à jouer pour suivre et mesurer les conséquences à long terme de l'utilisation de robots sociaux.** Maintenant que ces derniers sont utilisés à grande échelle, il s'agit d'en mesurer par des études interdisciplinaires l'incidence sur la cognition, le psychisme, le comportement des utilisateurs, le rapport de ces derniers à autrui et au monde, les conditions de la convivialité⁵³ entre humains et robots sociaux et entre les humains accompagnés de robots ; il s'agit de construire ainsi, en toute indépendance, un socle de savoirs nécessaires pour faire face aux enjeux liés à l'usage de ces outils et pour en anticiper les éventuels effets sociaux néfastes.

C'est pourquoi le COMETS appelle la recherche publique à **la vigilance et formule les recommandations suivantes.**

IV. RECOMMANDATIONS

Le COMETS recommande aux personnels de recherche (tout particulièrement en informatique et en robotique), aux sociétés savantes et aux organismes de recherche de :

- 1) **prévoir des formations aux enjeux éthiques** dans les cursus scientifiques et techniques concernés, d'une part, pour les personnels de recherche d'autre part ; ces formations devront conduire ces derniers à se familiariser davantage avec la littérature internationale consacrée aux questions éthiques évoquées dans cet avis ;
- 2) **s'interroger sur les finalités de la recherche**, des applications et des choix de conception, comme sur les avantages et les inconvénients à donner aux robots une forme ou un comportement humanoïde ou encore des capacités à capter et à simuler les émotions ;
- 3) mener, dans des situations et contextes réalistes, des **études scientifiques de long terme et à grande échelle** (suivi d'un large éventail d'individus) **sur : a/ les interactions entre humains et robots sociaux** pour comprendre plus finement comment les premiers perçoivent les seconds, quel degré de rationalité et de confiance ils leur prêtent, quelles relations ils établissent avec eux, quels effets bénéfiques ou négatifs en découlent,

⁵¹ Voir l'avis n° 3 du CNPEN, précité.

⁵² On vise ici les questions relatives à l'attachement, encore trop peu considérées et examinées, même si une littérature apparaît depuis quelques années sur la question. En revanche, il n'est pas question des devoirs juridiques et éthiques auxquels se plient d'ores et déjà les personnels de recherche en ce qui concerne la protection des personnes qui se prêtent aux expérimentations avec les machines, ou du traitement des données (car pour améliorer les performances de leurs modèles de robots, les chercheurs s'appuient sur des corpus de données personnelles de plus en plus vastes et variés d'interactions entre humains).

⁵³ I. Illich, *Tools for Conviviality*, Harper and Row, 1973.

notamment en termes cognitifs et psycho-affectifs (attachement, impacts sur les états émotionnels, les décisions, l'autonomie d'action) ; **b/ les interactions et conditions de la convivialité entre les humains lorsqu'ils sont accompagnés de robots sociaux.** Maintenant que de nombreux robots sociaux sont sur le marché et largement utilisés dans des situations « naturelles », de telles études scientifiques sont à la fois possibles et nécessaires ;

- 4) renforcer à cet effet les **recherches interdisciplinaires** associant aux travaux en informatique, en robotique ou en sciences du comportement, du traitement du langage, des recherches en psychologie, neurosciences, linguistique, sociologie, droit, éthique, philosophie et anthropologie ;
- 5) dans le cadre d'un **observatoire**, collecte à grande échelle et à long terme les données relatives à l'utilisation de robots sociaux et la manière dont les utilisateurs se les approprient⁵⁴ ; il s'agit d'alimenter la recherche scientifique, de construire un socle de connaissances indépendantes permettant d'adapter l'innovation dans ce champ, d'éclairer les décideurs sur les enjeux d'information, de mise en garde des utilisateurs, voire de réglementation des développements et applications commerciales des robots sociaux.

Outre les responsabilités qui doivent peser sur les ingénieurs concepteurs et industriels fabricants de robots sociaux, les utilisateurs et les pouvoirs publics, le COMETS entend ainsi souligner la responsabilité éthique qui incombe au monde de la recherche, d'étudier sur le long terme les effets de ces objets de manière à en permettre un usage libre et responsable.

⁵⁴ Voir dans ce même sens le CNPEN, qui appelle à la mise en place d'une entité souveraine de recherche et de formation « IA sciences et société ». Divers observatoires existent d'ores et déjà en matière d'IA : celui de l'OCDE sur les politiques étatiques et les incidents d'IA ; le réseau OBVIA québécois (Observatoire international sur les impacts sociétaux de l'IA et du numérique) ; l'observatoire de l'IA de l'Université Paris 1 Panthéon-Sorbonne qui, dans une « approche réflexive et multidisciplinaire », vise à fédérer les actions menées par différentes équipes de recherche de cette université sur le développement de l'IA dans différents champs (dont l'informatique, les mathématiques, le droit, l'histoire, la géographie, la philosophie, l'économie, la gestion et les arts), afin d'explorer l'impact de l'IA sur la société. L'observatoire recommandé par le COMETS pourrait être envisagé dans le cadre de cette dernière structure ou s'appuyer sur elle.

V. PERSONNALITÉS CONSULTÉES

- David Cohen, professeur en neurosciences à Sorbonne Université, chef du service de psychiatrie de l'enfant et de l'adolescent de l'hôpital Pitié-Salpêtrière, Paris
- Paul Dumouchel, professeur à l'université du Québec et chercheur au CREA
- Alain Hui Bon Hoa, chef des opérations et des produits, United robotics group
- Dominique Lestel, maître de conférences en philosophie à l'ENS
- Jérôme Monceaux, PDG et fondateur d'Enchanted Tools
- Vanessa Nurock, professeure de philosophie à Université Côte d'Azur
- Lionel Obadia, professeur en anthropologie sociale et culturelle à l'université de Lyon II
- Magalie Ochs, maîtresse de conférences à Aix-Marseille Université, Laboratoire d'Informatique et Systèmes
- Pierre Philip, chef du service de médecine du sommeil à l'université de Bordeaux, responsable de l'axe numérique de l'unité de recherche Sanpsy
- Serge Tisseron, psychiatre, psychanalyste, professeur émérite à l'université Paris Cité
- Denis Vidal, anthropologue, maître de conférences à l'EHESS et directeur de recherche à l'IRD

VI. BIBLIOGRAPHIE

- D. Andler, *Intelligence artificielle, intelligence humaine: la double énigme*, Gallimard, 2023.
- Belga, « Europol alerte sur les abus possibles de ChatGPT par les cybercriminels, *L'avenir*, 27 mars 2023.
- V. Aubergé, « Le robot social est un selfie de l'homme moderne », *The conversation*, 6 octobre 2019 ; « Des robots vraiment sociaux ? », *Le Monde*, 9 novembre 2017.
- L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez et S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements", *Psychological science in the public interest*, 20(1), 1-68, 15 juillet 2019.
- P. Bisconti, *Hybrid Societies: Living with Social Robots*, Taylor and Francis, 2024.
- C. Breazeal, K. Dautenhahn et T. Kanda, "Social robotics", *Springer handbook of robotics, 1935-1972*, 2016.
- C. Boldeau-Blandin interviewée par M. Protais, « On s'attache à des robots de plus en plus parfaits, qui vont toujours dans notre sens », *L'ADN*, 14 avril 2023.
- J. Bowlby, *Attachment and loss (Vol. 1): Attachment*, Basic Books, 1969.
- J. Bowlby, "The Bowlby-Ainsworth attachment theory", *Behavioral and Brain Sciences*, 2(4):637-638, 1979.
- D. Boyd, *C'est compliqué, les vies numériques des adolescents*, C&F éditions, 2016.
- F. Carmagnat et E. Robson, « Qui a peur du Tamagotchi ? Etude des usages d'un jouet virtuel », *Réseaux Communication - Technologie - Société*, n° 92-93 (pp. 343-364), 1999.
- J. Cassell, « Socially interactive agents as peers », in "The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational », *Agents, Intelligent Virtual Agents, and Social Robotics, Volume 2: Interactivity, Platforms, Application* (pp. 331-366), juillet 2022.
- CERNA (Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique de l'alliance Allistene), Rapport n° 1, *Ethique de la recherche en robotique*, 2014.
- R. Chatellier, « Captation des émotions : comment vous le direz pourra être retenu contre vous... », *LINC*, avril 2018 (<https://linc.cnil.fr/fr/captation-des-emotions-comment-vous-le-direz-pourra-etre-retenu-contre-vous>).
- M. Chollet, S. Marsella et S. Scherer, "Training public speaking with virtual social interactions: effectiveness of real-time feedback and delayed feedback", *Journal on Multimodal User Interfaces* (pp.1-13), 2022.
- CNPEN, avis n° 3, *Agents conversationnels : enjeux d'éthique*, 15 septembre 2021.

- CNPEN, avis n° 7, *Systèmes d'intelligence artificielle générative : enjeux d'éthique*, 30 juin 2023.
- Commission européenne, DG Recherche et Innovation, secteur de l'éthique et de l'intégrité de la recherche, *Ethics by design and ethics of use approaches for Artificial Intelligence*, 25 novembre 2021.
- Conseil de l'Europe, *Recommandation sur les impacts des systèmes algorithmiques sur les droits de l'homme*, CM/Rec (2020)1 (p. 192), 8 avril 2020.
- A. Cordier, *Grandir informés : Les pratiques informationnelles des enfants, adolescents et jeunes adultes*, C&F Editions, 2023.
- T. Curiale, F. Acquatella, L. Gros, M. Cosquer et S. Tisseron, « L'anthropomorphisme, enjeu de performance pour les chatbots », *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, vol. XXVIII (pp. 101 – 123), 2022.
- B. Dainow et P. Brey, "Ethics by design and ethics of use approaches for artificial intelligence" *Commission européenne, DG Recherche et Innovation*, 2021.
- A. Damasio, *Les furtifs*, La Volte, 2019.
- A. Damasio, *Vallée du Silicium*, Seuil, 2024.
- K. Dautenhahn, « Socially intelligent robots: dimensions of human–robot interaction », *Philosophical transactions of the royal society B: Biological sciences*, 362(1480) (pp 679-704), 2007.
- D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer et L.-P. Morency, « SimSensei Kiosk: A virtual human interviewer for healthcare decision support », *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 1061-1068), mai 2014.
- L. Devillers, *Les robots émotionnels*, L'Observatoire, 2023.
- L. Devillers, « Les systèmes d'intelligence artificielle et le langage ; enjeux d'éthique », in « Tout est-il souhaitable en recherche », *Raison présente*, n° 228 (pp. 65-72), 2023.
- A. Dimberton, « Kaspar : utiliser l'humanoïde comme 'médiateur social' », *Silicon*, 9 juillet 2007.
- C. Dolbeau-Bandin, *Un robot contre Alzheimer ; approche sociologique de l'usage du robot Paro dans un service de gériatrie*, C&F Editions, collection Interventions, 2021.
- ETX, « Néo-luddisme : quand la vie connectée n'est pas du goût de tous », *site Les Numériques*, 25 décembre 2022.
- O. Ezratty, « Où en est l'IA émotionnelle », *Opinions libres, blog d'Olivier Ezratty*, 2 mars 2018 (<http://www.oezratty.net/wordpress/2018/ia-emotionnelle>).
- S. Frachet, « IA Medical lève 1,4 million d'euros pour diffuser son chatbot contre Alzheimer », *L'Usine digitale*, 2 juin 2023.
- J.-G. Ganascia, *Le mythe de la singularité*, Le Seuil, 2017.
- E. Heyselaar, N. Caruana, M. Shin, L. Schilbach et E. S. Cross, "Do we really interact with artificial agents as if they are human?", *Frontiers in Virtual Reality*, 2023.
- B. Hill, « Intelligence artificielle : « Qui, de l'humain ou de la machine, est le plus fort ? », *Le Monde*, 18 septembre 2023 ; l'article s'inspire de l'étude de B. Hill, « Taking the help or going alone : ChatGPT and class assignments », *HEC Paris Research Paper* juin 2023.
- I. Illich, *Tools for Conviviality*, Harper and Row, 1973.
- K. Ishiguro, *Klara and the Sun*, Faber et Faber, 2021; traduit en français par Anne Rabinovitch sous le titre *Klara et le soleil*, Gallimard, 2023.
- W. L. Johnson, "Constructing virtual role-play simulations", *Design Recommendations for Adaptive Intelligent Tutoring Systems: Authoring Tools*, 3 (pp 211-226), 2015.
- T. Kanda, T. Hirano, D. Eaton et H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial", *Human–Computer Interaction*, 19(1-2) (pp. 61-84), 2004.
- J. Klayman, "Varieties of confirmation bias", *Psychology of learning and motivation*, 32 (pp.385-418), 1995.
- N. C. Krämer, "Social Effects of Virtual Assistants. A Review of Empirical Results with Regard to Communication", *IVA 2008: 507-508*, 2008.
- T. Law, M. Chita-Tegmark, N. Rabb et M. Scheutz, "Examining attachment to robots: benefits, challenges, and alternatives", *ACM Transactions on Human-Robot Interaction (THRI)*, 11(4), 1-18, 2023.
- Le Dépendant (rédaction), « Perpignan : elle se suicide car son Tamagotchi est mort », *Le Dépendant*, 3 juin 2019.

- N. F. Lindemann, "The Ethics of 'Deathbots'", *Science and Engineering Ethics* 28 (6) (pp.1-15), 2022.
- J. Loh et W. Loh, "Introduction, Social Robotics and the Good Life", in J. Loh and W. Loh, *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds with Robots*, Transcript Verlag (pp. 7-22), 2022.
- V. Lucchese, « Un Japonais a épousé un hologramme », *site Usbek et Rika*, 13 Novembre 2018.
- M. Mori, « Bukimi no tani gensho », *Energy*, vol. 7, no 4, 1970, traduction dans M. Mori et K. F. MacDorman, « La Vallée de l'Étrange de Mori Masahiro : Importance et impact sur l'esthétique et la conception des robots », *e-Phaïstos, revue d'histoire des techniques*, vol. 7, no 2, 2019.
- C. Muller (dir.), *The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law*, Report of the Ad hoc Committee on Artificial Intelligence (CAHAI), *Council of Europe, CAHAI(2020) 06 fin, no 21*, 24 juin 2020.
- D. Nora, « Intelligence artificielle : Replika, le chatbot amoureux... et dangereux ? », *L'Obs*, 11 septembre 2023.
- A. Paiva, I. Leite, H. Boukricha et I. Wachsmuth, I., "Empathy in virtual agents and robots: A survey", *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3) (pp.1-40), 2017.
- X. Pan, M. Gillies et M. Slater, "Virtual character personality influences participant attitudes and behavior—an interview with a virtual human character about her social anxiety", *Frontiers in Robotics and AI*, 2, 1, 2015 (<https://www.alelo.com/>).
- E. Petit, « La relation empathique homme/robot : normale ou pathologique? », *Implications philosophiques*, 21 juillet 2021 (<http://www.implications-philosophiques.org/la-relation-empathique-homme-robot-normale-ou-pathologique/#:~:text=De%20«%20normale%20»%2C%20la%20relation,cœur%20même%20de%20ce%20processus>).
- P. Philip, L. Dupuy, C.-M. Morin, E. de Sevin, S. Bioulac, J. Taillard, F. Serre, M. Auriacombe et J.-A. Micoulaud-Franchi, « Smartphone-Based Virtual Agents to Help the General Population Concerned by Sleep Complaints During COVID-19 Confinement: A Feasibility Study », *J Med Internet Res*, 2020 (<https://www.wefight.co/fr/>).
- R. W. Picard, *Affective computing*, MIT press, 2000.
- A. Placani, "Anthropomorphism in AI: hype and fallacy", *AI and Ethics* (pp;1-8), 2024.
- R. Plutchik, *Evolutionary bases of empathy. In Empathy and Its Development*, N. Eisenberg et J. Strayer Editeurs, Cambridge University Press, 1987 (<https://books.openedition.org/editionsmsmh/13902?lang=fr>).
- S. Raffin, « Alzheimer : une intelligence artificielle peut détecter les signes dans la voix », *site Pourquoi docteur*, 17 avril 2023.
- B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people and places", *CSLI Publications, Stanford*, 1969.
- M. Rochefort, « L'ingénieur de Google qui voyait l'intelligence artificielle comme une personne a été licencié », *Siècle digital*, 25 juillet 2022.
- J. Rochfeld et C. Zolynski, « La valeur des émotions : quel régime pour le 'capitalisme mental' ? », in *Mélanges en l'honneur de P. Sirinelli*, Dalloz (pp. 749-770), 2022.
- N. Sauer, « Deadbots' can speak for you after your death. Is that ethical? », *The Conversation*, 9 mai 2022.
- K. R. Scherer, "Emotion", in *M. Hewstone et W. Stroebe éditeurs, Introduction to Social Psychology. A European perspective*, Oxford Blackwell (pp. 151-191), 2000.
- I. Sivignon, « Bblue Monday : découvrez 4 chatbots messenger pour éviter la déprime », *magazine Influenth*, 20 janvier 2020.
- M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments", *Philosophical Transactions of the Royal Society B* (pp. 3549–3557), 2009.
- S. Tisseron, *Le jour où mon robot m'aimera : vers l'empathie artificielle*, Albin Michel, 2015.
- S. Tisseron, *Petit traité de cyberpsychologie*, Le Pommier, 2019.
- C. Vincent, « Serge Tisseron : Les robots vont modifier la psychologie humaine », *Le Monde*, 12 juillet 2018.

- A. M. Von der Pütten, N. C. Krämer, J. Gratch et S.-H. Kang, "It doesn't matter what you are!" Explaining social effects of agents and avatars", *Computers in Human Behavior* 26 (6) (pp.1641-1650), 2010.
- Matthew Weaver, « AI chatbot 'encouraged' man who planned to kill queen, court told» *The Guardian*, 6 juillet 2023.
- M. West, R. Kraut et H. E. Chew, "I'd blush if I could: closing gender divides in digital skills through education", *UNESCO, Equals Skills Coalition*, 2019.
- K. Winkle, D. McMillan, M. Arnelid, K. Harrison, M. Balaam, E. Johnson et I. Leite, "Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical HRI", in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 72-82), mars 2023.
- J. C. M. Wong, M. J. S. Wan, L. Kroneman, T. A. Kato, W. Lo, P. W. C. Wong et G. H. Chan, « Hikikomori Phenomenon in East Asia: Regional Perspectives, Challenges, and Opportunities for Social Health Agencies », *Frontiers Psychiatry Sec. Child and Adolescent Psychiatry Volume 10*, 23 juillet 2019.
- T. Xie et I. Pentina, "Attachment theory as a framework to understand relationships with social chatbots: a case study of Replika", *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022.
- A. Zhang et P. Patrick, "Tools or peers? Impacts of anthropomorphism level and social role on emotional attachment and disclosure tendency towards intelligent agents", *Computers in Human Behavior*, 138, 107415, 2022 (<https://doi.org/10.1016/j.chb.2022.107415>).

VII. LISTE DES ACRONYMES

CEA : Commissariat à l'énergie atomique et aux énergies alternatives

CERNA : Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique de l'alliance Allistene

CREA : Centre de Recherche en épistémologie appliquée de l'Ecole Polytechnique

CNPEN : Comité national pilote d'éthique du numérique

CNRS : Centre national de la recherche scientifique

COMETS : comité d'éthique du CNRS

EHESS : École des hautes études en sciences sociales

ENS : Ecole normale supérieure

FAQ : Foire aux questions

GPT : *Generative Pre-trained Transformer*

IA : intelligence artificielle

INRIA : Institut national de recherche en sciences et technologies du numérique

IRD : Institut de Recherche pour le Développement

LaMDA : *Language model dialogue application*

OBVIA : Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique

OCDE : Organisation de coopération et de développement économiques

OMS : Organisation mondiale de la santé

SNCF : Société nationale des chemins de fer français