



Disponible en ligne sur

ScienceDirect
www.sciencedirect.com

Elsevier Masson France

EM|consulte
www.em-consulte.com



VIE DE L'ACADÉMIE

Les robots sociaux et affectifs en santé[☆]

Social and emotional robots in healthcare

Un chatbot est un agent conversationnel capable d'échanges langagiers (à l'oral ou à l'écrit) dans un dialogue avec un utilisateur. Il est souvent intégré dans un système numérique : un smartphone, une interface sur un écran d'ordinateur, une enceinte vocale ou un robot. 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 [1] et aux modèles de langage pré-entraînés en libreaccès 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, les utilisateurs ont tendance de façon naturelle à les anthropomorphiser, c'est à dire à leur donner des capacités humaines. Le perfectionnement de ces technologies tend à brouiller la frontière perçue entre les machines et les êtres humains. 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 mais aussi aux techniques de personnalisation et de traitement des émotions. 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 transformeurs apprennent à partir de vastes corpus linguistiques les régularités les plus saillantes, sans être influencés par l'ordre des mots. Cette technique a été développée à partir de 2017 et depuis la mise à disposition du modèle de langage BERT (Bidirectional Encoder Representations from Transformers) par Google, de nombreuses avancées scientifiques ont vu le jour. Le premier agent conversationnel de l'histoire de l'informatique est le programme ELIZA de Joseph Weizenbaum (MIT, 1966) qui est aussi l'un des premiers leurres conversationnels. ELIZA simule un dialogue écrit avec un psychologue rogerien 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.

Les chatbots, vocaux ou écrits, sont déjà déployés dans le domaine de la santé : ils sont utilisés pour le diagnostic, la surveillance ou encore l'assistance aux patients. Des tâches de soin répétitives sont de plus en plus déléguées aux chatbots, par exemple l'information du patient avant et après une opération ou l'éducation et le suivi de patients diabétiques [2]. L'usage des chatbots en psychiatrie ouvre également à de nombreuses questions éthiques et de robustesse de ces systèmes.

Thaler et Sunstein [3] en 2008 appellent nudge, une stratégie de « coup de pouce » consistant à modifier subtilement le comportement de la personne. Les agents conversationnels peuvent détecter le comportement émotionnel d'une personne et adapter une stratégie de nudge. La synthèse émotionnelle d'un chatbot peut également servir pour influencer. Les « nudges numériques », tels que présentés par les chatbots et robots sociaux et affectifs sont définis comme des suggestions ou des manipulations manifestes ou subliminales (cachées) destinées à influencer le comportement ou les émotions d'un utilisateur [4,5]. L'utilisation du nudge peut être appropriée dans des situations telles que les soins en santé comme le traitement des addictions (alcoolisme, toxicomanie). Si cet aspect peut être fondamental pour des applications médicales, les chatbots doivent-ils être conçus pour inciter les gens à modifier leurs opinions et leurs actions pour le bénéfice personnel de l'utilisateur

[☆] Résumé d'une communication présentée lors de la journée dédiée « Médecine et intelligence artificielle » du mardi 3 mai 2022 à l'occasion de la parution de l'ouvrage : « Nordlinger B, Villani C, de Fresnoye O Dir. *Médecine et intelligence artificielle*. Paris : CNRS éditions, 2022 ».

et/ou pour quelqu'un d'autre ? Quelle transparence et audit a-t-on de ces systèmes ? Quelles tensions éthiques sont nécessaires à comprendre ?

Les nudges renforcés par l'IA (c'est-à-dire, tous les types de nudges numériques utilisant des mécanismes d'IA comme déductions statistiques du comportement des utilisateurs) soulèvent de nouvelles préoccupations éthiques. Les mécanismes de nudging peuvent intervenir à un niveau de granularité très fin (par exemple, l'adaptation du ton de la voix du chatbot sur la base de variantes d'apprentissage en ligne de l'apprentissage automatique). Ces processus peuvent générer un certain nombre de risques éthiques pour les individus, les groupes ou la société. Leur diffusion massive sur une longue période auprès d'un groupe spécifique de personnes peut avoir des implications sociales et économiques à grande échelle.

Nous sommes dans une phase de transition entre l'homme et la machine et dans l'émergence de relations humain-machine plus hybrides [6]. L'architecture de choix de la machine peut être modifiée pour nous influencer. Le nudge est une approche paternaliste puisqu'un groupe de personnes sait ce qui est bien pour nous. Il est donc important d'auditer ces systèmes pour vérifier qui a décidé et quelle stratégie est en place. Plus de transparence et d'explication des systèmes est nécessaire. Dans la majorité des cas, les chatbots actuels répondent selon des stratégies prédéterminées par leurs concepteurs ce qui est en train de changer avec l'utilisation de technologies comme les transformateurs. Ces nouvelles approches augmentent la difficulté de transparence des systèmes et risquent de renforcer les stéréotypes présents dans nos données. Etablir une norme devrait permettre de définir des nudges typiques déjà actuellement utilisés ou ceux qui pourraient être créés. Nous avons rendu un avis en 2021 sur les enjeux éthiques liés aux chatbots [7] au CNPEN, Comité National Pilote d'Éthique du Numérique. Des analyses systématiques et un renforcement des recherches sont nécessaires pour examiner l'éthique de la conception de ces systèmes pour guider les êtres humains avant leur déploiement [8].

Déclaration de liens d'intérêts

L'auteur déclare ne pas avoir de liens d'intérêts.

Références

- [1] Collins E, Ghahramani Z. LaMDA: our breakthrough conversation technology; 2021 [En ligne] Disponible sur : <https://blog.google/technology/ai/lamda> (consulté le 2/05/2022).
- [2] Klonoff DC, Kerr D. Digital Diabetes Communication: There's an App for That. *J Diabetes Sci Technol* 2016;10:1003–5.
- [3] Thaler R, Sunstein CR. *Nudge: improving decision about health, wealth and happiness*. New Haven, CT: Yale University Press; 2008.
- [4] Devillers L. *Les robots émotionnels*. Paris: L'Observatoire; 2020.
- [5] IEEE Affective Computing; 2018 [En ligne] Disponible sur : https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e_affective_computing.pdf (consulté le 2/05/2022).
- [6] Rahwan I, et al. *Machine behaviour*. *Nature* 2019;568:477–86.
- [7] Comité consultative nationale d'éthique. *Agents conversationnels : enjeux d'éthique*. [En ligne] Disponible sur : <https://www.ccne-ethique.fr/fr/actualites/cnpen-les-enjeux-ethiques-des-agents-conversationnels> (consulté le 2/05/2022).
- [8] Devillers L, Fogelman-Soulié F, Baeza-Yates R. *AI & Human Values, Reflections on Artificial Intelligence for Humanity*. Springer; 2021. p. 76–89.

L. Devillers
CNRS-LISN (Saclay), 508, rue John-von-Neumann, 91405
Orsay cedex, France

*Auteur Correspondant. LISN (ex LIMSI) CNRS- Campus
universitaire bât 507 Rue du Belvédère F - 91405 Orsay
cedex

Adresse e-mail : laurence.devillers@limsi.fr

Reçu le 2 mai 2022

Accepté le 2 mai 2022

Disponible sur Internet le 26 juillet 2022