

Un statut moral pour les robots ?

Le détour par l'interaction

Manuel Rebuschi

AHP-PReST, Université de Lorraine
manuel.rebuschi@univ-lorraine.fr

LES ROBOTS SOCIAUX ne sont pas des personnes comme les autres. On pourrait même considérer que ce ne sont pas des personnes du tout mais ce serait aller un peu vite en besogne. En effet, comme cela a été observé empiriquement au sein du groupe Psyphine (Becker et al., 2019 ; André & Boniface, 2021) mais également dans de nombreuses autres études (cf. une revue de la littérature pluridisciplinaire in (Spatola, 2019)), les humains ont un biais anthropomorphique dans leurs interactions comme dans leur observation des interactions avec les robots sociaux. Les humains semblent ainsi avoir une fâcheuse tendance à projeter une intériorité sur de drôles d'objets pour peu que ceux-ci soient pris dans des interactions, et cela alors même que personne n'y croit (Rebuschi, 2022).

Par son côté systématique, ce clivage entre pratique et croyances qui pourrait ressembler à un jeu de dupes cache certainement quelque chose de profond. Si l'interaction humain-robot joue un rôle tellement important dans l'attribution, même nuancée ou estompée, d'une intériorité donc d'une potentielle personnalité à un robot, c'est qu'elle offre peut-être un bon point de départ pour l'analyse. De façon générale on peut considérer qu'une entité a un statut moral quand elle peut nous donner des raisons d'agir d'elle-même et pour elle-même (Kamm, 2007). Et si le statut moral d'une entité, plutôt que de se fonder sur l'ontologie autrement dit sur la nature de cette entité, venait s'ancrer sur la qualité de nos interactions et plus généralement de nos relations avec elle ?

Cette stratégie est suggérée par les approches relationnelles de la personnalité, relativement étrangères à la culture occidentale mais qui ont été repérées dans d'autres cultures (Bird-David, 1999 ; Fowler, 2016). Ces approches ont conduit des philosophes, notamment sud-africains, à promouvoir le relationnisme moral, également connu sous l'appellation ubuntu, suivant lequel le statut moral doit être fonction de nos interactions et relations sociales plutôt que de l'ontologie individuelle (Metz, 2012 ; Ewuoso & Hall, 2019). Cette conception est, de manière attendue, mobilisée dans le cadre de l'éthique de l'environnement à l'appui de l'attribution d'un statut moral aux animaux non-humains (May, 2014), mais elle est aussi explicitement défendue pour l'attribution d'un statut moral aux robots (Coeckelbergh, 2014 ; Gunkel, 2022 ; Jecker et al., 2022).

Le relationnisme moral n'est pas la seule conception possible ni même la seule effectivement défendue au sujet des robots (Gordon & Gunkel, 2022). On peut cependant argumenter que parmi les conceptions « libérales » ou inclusives vis-à-vis des robots, elle est la plus naturellement compatible avec le clivage entre pratique et croyances mentionné plus haut. Encore faut-il s'assurer que l'attribution d'un statut moral en s'appuyant sur la pratique n'est pas purement subjective et gratuite, autrement dit qu'à défaut de coïncider avec nos croyances elle répond à des critères normatifs.

Suivant une analyse classique et standard, une interaction sociale authentique implique nécessairement de l'intentionnalité. Une difficulté surgit avec l'interaction humain-robot du fait qu'elle est à demi-fictionnelle (puisqu'on simule, qu'on fait comme si, sans y croire) et à demi-sérieuse (puisque l'interaction se produit dans un environnement physique, social et culturel réel) : elle a le caractère d'une « fiction autoritaire », étant en cela analogue

aux fictions juridiques (Rebuschi 2023, 2024). S'agit-il d'une interaction authentique ? Tout dépend alors du statut de l'intentionnalité.

Une approche perspectiviste donc antiréaliste de l'intentionnalité, inspirée de la stratégie de l'interprète (intentional stance) de Dennett (1990), permet d'aligner au moins certaines des interactions (à demi-fictionnelles) humain-robot sur les interactions (sérieuses) humain-humain : les premières ne sont pas moins authentiques que les secondes puisque dans les deux cas l'intentionnalité est projetée par l'observateur-interprète (qui peut être participant à l'interaction). L'intentionnalité projetée sur l'interaction l'est indépendamment de la nature ou des propriétés des entités qui interagissent ; elle peut l'être également indépendamment de, voire préalablement à la projection d'intentionnalité à ces entités elles-mêmes.

Selon cette stratégie, l'intentionnalité n'est pas projetée gratuitement mais conformément à certains critères normatifs rationnels, épistémiques, heuristiques ou autres. Il est ainsi possible d'être antiréaliste au sujet de l'intentionnalité sans pour autant être subjectiviste ou relativiste. Or parmi les normes qui viennent contribuer à qualifier l'interaction, on peut naturellement inclure des normes éthiques. L'approche centrée sur l'interaction, et plus généralement sur les relations entretenues entre entités, s'étend alors aisément de l'attribution d'intentionnalité à celle d'un statut moral, tout en accommodant les situations, complexes par leur mélange de fiction et de sérieux, de l'interaction humain-robot.

RÉFÉRENCES

- André, V., Boniface, Y. (2021). Des balbutiements à DECIDE : les tribulations de la lampe Psyphine. In *Collectif Psyphine* (eds.). *Que prêtons-nous aux machines ? Approches interdisciplinaires des interactions homme-robot*. PUN – Éditions Universitaire de Lorraine, 191-213.
- Becker, J., André, V., Dutech, A. (2019). Qualcom : une expérience sur la qualification des comportements d'une lampe robotique. *Techniques & Culture, Varia* [en ligne]. <http://journals.openedition.org/tc/10771>
- Bird-David, N. (1999). "Animism" revisited: Personhood, environment and relational epistemology. *Current Anthropology* n° 40 (Supplement), 67-91.
- Coeckelbergh, M. (2014). The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics. *Philosophy and Technology* 27(1), 61–77.
- Dennett, D. (1990). *La stratégie de l'interprète*, Paris, Gallimard (The Intentional Stance, The MIT Press, Cambridge, Mass., 1987).
- Ewuoso, C., Hall, S. (2019). Core aspects of ubuntu: A systematic review. *S Afr J Bioethics Law* 12(2), 93-103.
- Fowler, C. (2018). Relational Personhood Revisited. *Cambridge Archaeological Journal*, 26(3), 397–412.
- Gordon, J.-S., Gunkel, D.J., (2022), Moral Status and Intelligent Robots. *The Southern Journal of Philosophy* 60(1), 88-117.
- Gunkel, D.J. (2022). The Relational Turn: Thinking Robots Otherwise. In J. Loh & W. Loh (eds.), *Social Robotics and the Good Life*. Transcript Verlag, 54-76.
- Jecker, N.S., Atiure, C.A., Ajei, M.O. (2022). The Moral Standing of Social Robots: Untapped Insights from Africa. *Philosophy & Technology* 35, article number 34.
- Kamm, F. (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*, Oxford, Oxford University Press.
- May, T. (2014). Moral Individualism, Moral Relationalism, and Obligations to Non-human Animals. *Journal of Applied Philosophy* 31(2), 155-168.
- Metz, T. (2012). An African Theory of Moral Status: A Relational Alternative to Individualism and Holism. *Ethic Theory Moral Practice* 15(3), 387–402.
- Rebuschi, M. (2022). Drôles d'objets ou drôles d'interactions ? *Actes de Drôles d'objets: un nouvel art de faire*, Oct 2021, La Rochelle, France. [doi : 10.5281/zenodo.6059665]
- Rebuschi, M. (2023). Trois manières d'être une personne (pour un robot). *Actes de Drôles d'objets: un nouvel art de faire*, Mai 2023, Nancy, France.
- Rebuschi, M. (2024). Fictionnalisme vs. Animisme juridique. In J. Sohnle, Ch. Bouriau (dir.), *Éthique environnementale pour*

juristes / Environmental ethics for lawyers / Ética medioambiental para jurista. Mare & Martin, 303-320.

Spatola, N. (2019). L'interaction Homme-Robot, de l'anthropomorphisme à l'humanisation. *L'année psychologique* 119, 515-563.