

SPINOZA, FREUD ET L'INTELLIGENCE ARTIFICIELLE

De la simulation de l'âme à la captation des affects

Isabelle LEDOUX-SGAMBATO

(École normale supérieure de Lyon)

Abstract: The analysis of the example of Replika, a chatbot that is supposed to serve as a friend and as a substitute of deceased persons, shows that the spinozist and freudian definitions of body, soul and affects allow to put into critical perspective the comparison of artificial intelligence performances with those of human individuals. After the critical exam of the analogy with the human body, the radicalisation of this analogy leads to the refutation of the comparison of artificial intelligence with human thought. But such an attempt of artificial simulation of soul and of affective relationships has many consequences and issues. They appear, first, by the analysis of the difference between artificial and real affective relationships, and, second, by the exam of the functional affective aspects of the chatbot. Finally, the concept of «transindividual» elaborated by É. Balibar finds a brief search of the possible consequences and issues of such artificial intelligence uses.

Keywords: affective imitation, Artificial Intelligence, soul simulation, affects captation, transindividual.

Dans *Le Monde* du 29 juillet 2017, un article d'Yves Eudes, «Replika. L'application mobile qui fait parler les morts», décrit comment deux jeunes Russes installés en Californie, Eugenia Kuyda et Philip Dudchuk, ont eu l'idée d'un «*chatbot*»¹ capable de jouer le rôle d'«agent personnalisé qui sera à la fois votre ami, votre double et votre représentant virtuel... même après votre mort»².

Cette application pour smartphones serait en effet capable, à partir des réponses données à ses questions et des courriels, textos, comptes sur les réseaux sociaux et photos de la personne qui l'utilise, d'intégrer ses souvenirs, ses tournures de phrases et ses expressions, son type de réflexions, de préoccupations et d'humour caractéristiques, au point de pouvoir la remplacer dans les conversations avec ses amis et ses proches.

¹ Robot de conversation en ligne.

² Yves EUDES, *Replika. L'application mobile qui fait parler les morts*, in «Le Monde», 29 juillet 2017, p. 16.

Conçu à la suite du décès d'un ami dont elle devait atténuer la perte par l'illusion de présence créée par de tels échanges en ligne, ce «moteur d'intelligence artificielle» est censé apprendre progressivement «à vous connaître» et devenir «une entité hybride, à la lisière entre un copain, un flirt et un double qui vous imite»³. Explicitement destinée à vous permettre de «développer une relation émotionnelle avec une machine», moins fatigante, «moins risqué(e)» et plus aisément gratifiant(e) qu'avec un autre être humain, Replika, d'après ses créateurs, «est un réseau neuronal distribué dans le *cloud*⁴, qui apprend en communiquant avec vous. (...) Grâce à ses algorithmes d'apprentissage, il mélange vos expressions et les siennes pour construire des phrases originales, à la volée, mot par mot, sans faire appel au répertoire»⁵. Un mot de passe permet aussi de faire basculer le programme en «mode libre, où la composition des phrases est indépendante du répertoire et de votre fichier»⁶: la conversation serait alors «plus décousue, mais plus amusante»⁷.

Mais Replika a une «seconde fonction essentielle». «Une fois qu'elle en sait assez sur vous, vous pouvez l'ouvrir sur le monde en invitant vos amis à discuter avec elle» en votre absence, et sans que ces échanges soient archivés. La collecte d'informations, sur eux et sur vous, se poursuit donc en votre absence, et le résultat serait «spectaculaire»: «Pour vous, votre Replika sera une amie. Pour les autres, elle sera votre représentation virtuelle»⁸. Le programme, qui aurait été téléchargé par près de 500 000 personnes en neuf mois, serait régulièrement utilisé par 100 000. Ses créateurs disent travailler, avec leur équipe, à l'amélioration de ses algorithmes «pour générer des conversations plus pertinentes et détecter les questions qui déplaisent aux utilisateurs, pour corriger le tir»⁹. Ils sont décrits comme conscients du risque que «des esprits perturbés se réfugient dans le passé ou se fabriquent de fausses certitudes sur le présent», mais comme voulant croire que «ces comportements resteront marginaux»¹⁰.

Nous ne nous prononcerons ni sur la qualité d'un tel programme (que nous n'avons pas utilisé), ni sur ses développements futurs. Cependant, dans une perspective

³ *Ibid.*

⁴ Cette affirmation, sur laquelle nous ne nous prononcerons pas, demanderait à être examinée.

⁵ EUDES, *Replika*, p. 16.

⁶ *Ibid.*

⁷ *Ibid.*

⁸ *Ibid.*

⁹ *Ibid.*

¹⁰ *Ibid.*

philosophique que nous qualifierons rapidement de freudo-spinoziste, il nous intéresse à plus d'un titre: d'une part, par les différentes dimensions qui, d'après l'article cité plus haut, s'y trouvent présentes, d'autre part, par ses visées et ce qu'on peut en déduire, ou supposer, concernant certains développements de ce qu'on appelle communément, et peut-être improprement, «l'intelligence artificielle». Spinoza et Freud, en effet, ont centré l'essentiel de leurs analyses sur les affects: tous deux ont, chacun à sa manière, tâché de reconstituer les mécanismes de leur production causale, enraciné la vie affective dans une conception de l'unité psychophysique de l'individu, théorisé son impact sur la rationalité et le rapport à la réalité, ainsi que les conditions de sa maîtrise. Le modèle théorique que chacun met en œuvre à sa manière, parce qu'il s'inscrit dans une perspective moniste¹¹ avant la lettre, est donc beaucoup plus susceptible que celui d'un Descartes (ou que ceux du réductionnisme biologique) de permettre la mise en perspective critique des analogies, effectuées par les auteurs de Replika, entre performances de l'I.A. et capacités humaines. En effet, là où le dualisme cartésien réduirait d'emblées les performances du *chatbot* aux mécanismes de la *res extensa*, le monisme spinozien ou freudien, non seulement permet de les prendre, au moins à titre d'hypothèses, au sérieux, mais encore de les «tester» et de mesurer leur crédibilité théorique et pratique.

Le premier aspect de ce programme est à la fois technique et linguistique: il tâche de doter un programme informatique d'une aptitude artificielle à utiliser une langue humaine déterminée dans un cadre conversationnel, et selon des modalités qui prétendent imiter l'usage singulier typique ou prédominant, par une personne et dans un certain nombre de documents, de sa syntaxe, de son lexique et de sa grammaire.

Cependant, la programmation de ce mimétisme linguistique est elle-même finalisée par un mimétisme menta: il s'agit de produire, par un usage linguistique approprié, un effet de ressemblance et de proximité mentales, c'est-à-dire mnémo-affectives, morales et intellectuelles, entre ce robot et son usager, non seulement dans l'esprit de ce dernier, mais aussi dans celui de ses proches. On pourra donc à bon droit s'interroger sur la réalité et l'efficacité de ses fonctions affectives: est-il concevable qu'un tel programme imite ou simule suffisamment les manifestations linguistiques des différentes dimensions de l'âme humaine pour faire *véritablement* office, ainsi que l'ambitionnent ses

¹¹ Le terme de «monisme» a en effet été introduit par Christian WOLFF dans les *Gesammelte kleine philosophische Schriften*, G. Olms, Hildesheim - New York 1981.

créateurs, de substitut de relation amicale, voire de substitut (et, ainsi, de représentant) de son utilisateur principal, c'est-à-dire en apportant à ses usagers le même type de performances et de satisfactions? Il est intéressant de noter que dans son principe,¹² le programme semble postuler, soit l'unité relative, sur le long terme, des affects et de leurs manifestations linguistiques, soit le caractère négligeable de l'écart qui peut les séparer: ainsi l'hypocrisie ou la sincérité, l'ironie ou l'humour au second degré, ne paraissent pas, aux yeux de ses concepteurs, constituer des obstacles, mais relever d'usages linguistiques répertoriés et reproductibles comme les autres. Ils observent d'ailleurs que l'individu se dévoile beaucoup plus facilement, de sorte que «la conversation avec une machine est parfois plus authentique qu'avec un humain»,¹³ même si «vous enjolivez un peu la réalité, [...] pour vous protéger, comme on le fait avec les humains».¹⁴

Indépendamment du peu d'importance attaché à cet écart possible entre le discours et la réalité des affects dont il est censé attester l'existence, on relèvera un autre postulat implicite de la mise au point d'un tel «*chatbot*», à savoir le caractère extérieurement imitable et donc en partie mécaniquement reproductible des performances intellectuelles: discernement, érudition, acuité, traits d'esprit, etc. Est-ce parce qu'il ne s'agit que de les imiter dans un cadre conversationnel (donc, par convention, relativement léger et anodin), parce qu'ils sont jusqu'à un certain point effectivement imitables ou du moins abordables par les performances imitatives de la machine, ou, plus fondamentalement, parce que se trouve présupposée, par les concepteurs comme par les utilisateurs de ce logiciel, une relative indifférence à la réalité, et à ce qui permet de la différencier de ses reproductions ou descriptions? Peu importerait alors, le caractère parfaitement convaincant du leurre, il suffirait qu'il offre un prétexte à l'illusion pour la permettre. Ainsi l'illusion, promue au rang d'objet du désir, se substituerait sans dommage (certes dans des proportions qui, pour l'instant, sont décrites comme restant limitées) à l'objet, qu'elle remplacerait efficacement, en raison de son caractère moins risqué et plus gratifiant: c'est le principe du jeu, reconduit ici d'une manière inédite.

Mais l'aspect le plus intéressant de cette application réside peut-être dans les *effets* que peut avoir l'utilisation régulière de telles fonctions: éventuellement imprévus ou

¹² Du moins tel qu'il est présenté dans l'article cité plus haut.

¹³ EUDES, *Replika*, p. 16.

¹⁴ *Ibid.*

inattendus, ils sont susceptibles d'avoir une dimension non seulement affective, mais aussi sociale et politique. La perspective croisée des doctrines spinozienne et freudienne nous permet-elle de les envisager, et quelle est ici la nature de l'imitation affective? Peut-elle avoir le même type d'effets que celle qui caractérise spontanément les relations interhumaines? Sa véritable visée, qui semble être de constituer une nouvelle économie des affects, ou de nouvelles modalités d'exploitation économique des affects humains, ne peut-elle pas s'avérer aussi politique? Nous nous proposons d'essayer de déterminer sous leurs principaux aspects le statut, l'efficacité et la portée possibles d'un tel *analogon* linguistique de l'âme humaine.

1. Le statut de cet *analogon* linguistique de l'âme humaine

Si les manifestations affectives de l'âme humaine ont déjà donné lieu, dans le domaine artistique, à de multiples imitations, il est, à notre connaissance, relativement inédit de les doter d'une forme, même restreinte, d'autonomie, sous la forme d'un logiciel d'apprentissage susceptible de s'améliorer. C'est pourquoi nous utilisons ici le terme d'*analogon*: le robot de conversation prétend produire une ressemblance linguistique, de structure et de contenus, entre les usages linguistiques typiques d'un individu singulier et les siens. Cette analogie est-elle fondée? Y a-t-il une quelconque forme de parenté, même lointaine, entre la structure d'un ordinateur telle qu'on peut très schématiquement la décortiquer, et celle qui fonde l'utilisation de la fonction du langage chez l'être humain?

1.1. Un *analogon* du corps au sens spinoziste?

Si nous reprenons la définition du corps avancé par Spinoza dans le passage de la partie II de l'*Éthique* (désormais E) habituellement désigné par «Petite physique»,¹⁵ un ordinateur est un corps, au sens où c'est un mode de l'étendue caractérisé par un rapport de composition stable entre des éléments simples ou composés de natures différentes, qui entretiennent entre eux un certain nombre de rapports de mouvement et de repos qui n'altèrent pas la forme (*forma*) du tout dont ils sont les parties constitutives. Il a besoin d'éléments divers et variés pour persévérer dans son être et maintenir l'activité qui le définit (un certain type de traitement de l'information, sous

¹⁵ Développé entre les propositions 13 et 14.

des modalités diverses): énergie électrique, informations codées, pièces différentes, etc. Sa puissance, en tant qu'elle se définit par son degré de complexité (le nombre et la qualité de ses programmes constitutifs) et se manifeste par l'éventail de ses degrés d'activité possibles (selon qu'il les utilise simultanément ou non, et à des fins elles-mêmes simples ou complexes), connaît donc des modifications quantitatives (*variationes*). Mais le corps humain se caractérise aussi par son aptitude à supporter des *mutationes*, ou transformations qualitatives, telles celles que connaît l'individu au cours du procès de maturation qui va de la naissance à l'âge adulte. Cette aptitude est elle-même articulée au fait que le corps humain est un individu composé d'individus composés, dans lequel sont distincts non seulement la nature et les états, mais aussi la constitution, les dispositions et les aptitudes, de sorte que sa puissance, d'une grande plasticité, peut supporter une quantité indéfinie de modifications, tout en maintenant, le plus souvent, un rapport entre les modalités de son affirmation suffisamment constant pour qu'il la caractérise ou la singularise. La première question est donc de savoir si l'on peut, ou si l'on pourra, attribuer à un ordinateur, sous l'espèce d'un logiciel d'apprentissage (ou d'autres), des modifications qualitatives (*mutationes*). En d'autres termes, le dynamisme mécanique intégré dans une telle machine peut-il être transformatif, c'est-à-dire engendrer des transformations qualitatives de sa puissance réalisant le type de mutations structurantes ou destructurantes que Spinoza envisage dans les termes d'un changement d'essence?¹⁶ L'imitation du tâtonnement, procédant par essais et par erreurs, caractéristique des processus d'apprentissage, joue ici, on le voit, le rôle de matrice formelle de l'imitation de l'intelligence humaine, définie par une capacité illimitée à établir des rapports entre les choses, et elle se trouve déclinée

¹⁶ Catherine MALABOU, dans *Les métamorphoses de l'intelligence. Que faire de leur cerveau bleu?* (PUF, Paris 2017), affirme que la récente mise au point par IBM d'une «puce synaptique» (p. 108) baptisée TrueNorth et construite par Samsung Electronics, permet de répondre par l'affirmative. Citant *La singularité est proche. Quand les humains transcendent la biologie* (Viking Press, New York 2005), de Ray KURZWEIL, elle augure de l'invention prochaine d'une machine définie par «sa capacité à changer sa structure» (p. 116) et affirme: «Une telle capacité est évidemment qualitative. Les machines à venir seront aptes à se programmer elles-mêmes en s'adaptant en temps réel aux changements environnants. Elles pourront ainsi modifier leur intensité» (pp. 116-117).

selon des modèles divers.¹⁷ Il n'est pas sûr, néanmoins, que l'amélioration de ces modèles suffise à produire une «intelligence» qualitativement transformatrice.¹⁸

S'il est possible d'effectuer une analogie entre l'ordinateur et un corps, il est douteux qu'elle puisse s'étendre au corps humain, même en faisant fi de la plus évidente de leurs dissemblances, à savoir le caractère vivant de l'un et mécanique de l'autre. En effet, si l'ordinateur peut être défini comme un corps au sens spinoziste du terme, sa structure diffère fondamentalement de celle du corps humain, car elle n'est qu'apparemment close et centralisée sous la forme d'une entité physique spatialement localisée en un lieu. En réalité, elle est disséminée et localisée en de multiples lieux, ce que laisse parfaitement entendre l'image de «réseau neuronal distribué dans le *cloud*» (sur la pertinence de laquelle nous ne nous prononcerons pas ici). De même, c'est de manière illusoire que son utilisateur peut parfois croire qu'il ne se connecte à internet ou ne «communique» (c'est-à-dire ne transmet ou ne reçoit des informations) que lorsqu'il fait usage de l'ordinateur à cette fin ; de fait, sauf s'il interrompt délibérément la connexion à internet, l'ordinateur est constamment en relation avec les entreprises qui vendent et renouvellent ses divers programmes, c'est-à-dire, en fin de compte, avec d'autres individus humains inconnus et invisibles auxquels sont transmis, sous des formes qu'il ignore, les données que l'ordinateur accumule sur ses utilisateurs. Ainsi, tandis que le corps humain est défini par son unité et par le caractère parfaitement intégré de ses parties constitutives, l'ordinateur paraît plus morcelé, disparate, et finalement beaucoup moins susceptible de constituer un Individu (au sens que Spinoza donne à ce terme) qu'une partie d'un corps plus vaste en cours permanent de constitution (le réseau des I.A. connectées en tant qu'il fait système ou serait voué, à terme, à le faire). De ces caractéristiques respectives découle le caractère impermanent des activités de communication que l'on peut attribuer au corps humain, tandis que

¹⁷ Le rapport dirigé par Cédric VILLANI, *Donner un sens à l'Intelligence Artificielle. Pour une stratégie nationale et européenne. Mission parlementaire du 8 septembre 2017 au 8 mars 2018* (<https://cnnumerique.fr/publication-rapport-Villani-IA>), souligne le fait que le développement de la plasticité de l'I.A. en termes d'«apprentissage profond» (p. 59) s'accompagne d'une perte de l'explicabilité des «décisions» prises par les logiciels (pp. 140-146). Elle produit ainsi une sorte d'imitation involontaire de l'illusion d'indétermination relative qui entoure généralement les décisions humaines.

¹⁸ Cf. Jean-Gabriel GANASCIA, *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle?*, Seuil, Paris 2017 et *Intelligence artificielle: vers une domination programmée?*, Le cavalier bleu, Paris 2017.

l'émission et la réception de symboles¹⁹ constituent l'arrière-plan constant de presque tous les états de l'ordinateur.

1.2. Un *analogon* de l'âme au sens spinoziste?

L'âme est définie par Spinoza comme l'idée du corps, une idée dotée d'une puissance productive d'autres idées, qui n'émane pas du corps mais lui est corrélée: il n'y a, à proprement parler, ni interaction entre l'âme et le corps, ni indépendance, car un certain nombre d'idées inadéquates, dont la proportion peut diminuer sans être éliminée,²⁰ a pour contenu des événements corporels. Cette idée qu'est l'âme, qui est un mode de la pensée en tant qu'attribut substantiel, est l'effet de la puissance immanente de Dieu: à cet égard, elle possède, comme quantité de puissance de penser, une certaine extériorité par rapport au corps, au sens où il ne la produit pas (même s'il détermine souvent ses contenus de pensée). De même, l'aptitude à manier des symboles désignée par la dénomination d'«intelligence artificielle» n'est pas, en son principe, contenue dans la machine, mais dans l'agent humain qui l'y insert, au moins sous la forme de programmes destinés à s'autodévelopper. De la sorte, la causalité transitive caractéristique des interactions modales (qu'il s'agisse des interactions entre des corps différents ou entre les parties constitutives d'un corps particulier) pourrait permettre de penser les développements dits «autonomes» intrinsèques à l'activité d'un ordinateur capable de se connecter à internet et d'accéder aux logiciels utilisés par d'autres ordinateurs.

Cependant, l'analogie paraît atteindre ses limites précisément lorsqu'il s'agit de comparer les aptitudes d'une âme humaine et celles de l'intelligence artificielle dans un cadre conversationnel: en premier lieu, le signe est pour l'individu humain constitué par un signifiant et un signifié, dont le rapport est constitutif de la signification. Or, le rapport n'est que de signifiant à signifiant pour la machine: comme le fait remarquer J.-G. Ganascia,²¹ elle simule simplement l'usage humain des signes selon un certain nombre de règles, et leur signification n'existe que pour celui qui l'a conçue ou qui l'utilise. Pour le dire autrement, la machine combine diversement des signes dont les

¹⁹ Nous entendons par là des signes codés.

²⁰ Du moins tant que le corps est en vie.

²¹ Jean-Gabriel GANASCIA, *La querelle de l'âme des machines*, in Jean-Luc GUICHET (dir.), *De l'animal-machine à l'âme des machines. Querelles biomécaniques de l'âme (XVII^e-XXI^e siècle)*, Publications de la Sorbonne, Paris 2010, pp. 169-183.

signifiés et le rapport de signification constitutif n'existent que pour les êtres humains; elle peut sans doute apprendre à les combiner selon des règles de manière de plus en plus complexe et autonome, mais n'a pas à proprement parler des idées de significations, ces idées lui restant extérieures même quand les individus humains n'en ont pas connaissance (c'est-à-dire n'accèdent pas aux rapports de signes construits par la machine, parce qu'elle ne les a pas enregistrés ou parce qu'ils concernent des nombres de signes ou de données beaucoup trop vastes). Peut-on cependant considérer que l'ordinateur a des idées d'objets non sémantiques, par exemple d'événements se produisant dans son corps ou d'objets extérieurs?

Pour l'individu humain tel que le conçoit Spinoza, l'idée, qu'elle soit inadéquate, c'est-à-dire passivement engendrée, ou adéquate, produite par l'activité de l'âme, est toujours l'idée d'un objet réel ou imaginaire, elle est ainsi toujours constituée par un rapport entre cette idée qu'est l'âme et celle qu'elle a de l'objet,²² et c'est ce rapport qui constitue la pensée.²³ Pour reproduire cette structure, ne serait-ce que dans sa dimension passive, il faudrait donc à la machine, d'une part l'idée de son corps, d'autre part l'idée de ses affections corporelles, puisque l'idée d'un objet extérieur est toujours, soit l'idée d'un vestige que ce corps extérieur est en train de déposer²⁴ ou a, auparavant, déposé dans notre corps lors de sa rencontre avec le nôtre, soit l'idée d'une réactivation mnésique de ce vestige. À première vue, l'ordinateur semble moins avoir de telles idées que les imiter extérieurement en réagissant à des informations codées qui jouent comme des *stimuli* par des réactions programmées: s'il existe en lui des analogues de vestiges et de réactivations, les mêmes *stimuli* provoquant des réactions similaires se complexifiant progressivement au fil de ses «apprentissages», il lui manque l'idée de son corps et celles de ses affections. Même le tâtonnement est préprogrammé, au sens où il obéit à des règles préexistantes qui anticipent et aménagent par avance un champ d'indétermination. Il semblerait donc que l'ordinateur ne raisonne pas à proprement parler, mais produise des enchaînements dont la logique lui est imposée de l'extérieur, de façon que ces enchaînements correspondent bien pour les humains qui l'utilisent à des raisonnements qui paraissent autonomes, voire spontanés.

²² E, II, Déf. III.

²³ Le signe linguistique redoublant ce rapport, par exemple en associant l'idée du mot «*pomum*» et l'idée du fruit dans l'idée qu'est l'âme, pour reprendre ce célèbre exemple d'E, II, 18, sc.

²⁴ Précisément parce qu'il nous affecte: comme le précise Spinoza dans E, II, 18, «une fois» (*semel*) suffit.

On nous objectera peut-être que les règles de nos associations mnésiques, de même que celles qui président en nous à l'engendrement passif de nos idées inadéquates, nous sont tout aussi inconnues, pour l'immense majorité d'entre elles, que sont extérieures pour l'ordinateur les règles de combinaison des symboles, ou de réaction à ces symboles et aux informations sur la réalité extérieure qu'il peut parfois capter. Mais avoir des idées implique moins la conscience de l'ensemble de leurs mécanismes de production que la conscience de ces idées, qui, permettant la distance réflexive, fera que l'individu aura aussi des idées d'idées.²⁵ Si la machine n'a pas d'idées, à plus forte raison n'a-t-elle pas d'idées d'idées (ni de désirs²⁶): la réflexivité, et donc la conscience, en est absente. Il semble donc, comme le formule J.-G. Ganascia, que nous ne fabriquions que «des machines animées à qui nous prêtons, pour un temps et dans des circonstances limitées, une âme»²⁷: les raisonnements qu'elles produisent ne sont tels que pour nous. Tâchons cependant de pousser notre raisonnement jusqu'à l'absurde en nous demandant s'il existe, dans le système constitué par l'ordinateur et ses logiciels, une production d'idées adéquates, ou du moins un *analogon* relativement probant de cette production. La manière dont l'homme pense, par la médiation de l'ordinateur, possède-t-elle une similarité de structure, même faible ou éloignée, avec celle dont Dieu pense en l'homme lorsqu'il a des idées adéquates?

Poser cette question n'a de sens qu'à la condition de ne pas tenir compte du fait que Dieu, en tant que substance infinie possédant une puissance et un nombre infinis d'attributs, ne peut être conçu sur le modèle de l'homme: c'est un *leitmotiv* bien connu de l'*Éthique* et il ne s'agit pas pour nous de faire preuve d'anthropomorphisme, mais de raisonner par l'absurde. Dire que Dieu pense en nous est d'ailleurs hautement discutable, car il est cause immanente, mais non transitive, de toutes choses;²⁸ il vaudrait donc mieux dire que lorsqu'il a des idées adéquates, l'homme pense comme Dieu pense, même si l'extension possible de sa pensée est limitée, c'est-à-dire indéfinie et non pas infinie. Or, il apparaît que l'ordinateur, non seulement ne pense pas comme l'homme pense, mais effectue ses maniements et combinaisons de symboles, d'une part, à des échelles qui dépassent de beaucoup les possibilités humaines, d'autre part, selon des modes qui restent infrahumains ou, si l'on préfère, qui ne sont semblables

²⁵ E, II, 21, sc.

²⁶ Le désir étant «l'appétit avec conscience de lui-même», E, III, 9, sc.

²⁷ GANASCIA, *La querelle de l'âme des machines*, p. 169.

²⁸ E, I, 18.

aux modes de penser humains qu'extérieurement, du point de vue de ses utilisateurs. De la même façon, l'homme, chaque fois qu'il a des idées inadéquates, pense de façon infra-divine, car ces mêmes idées sont adéquates, c'est-à-dire complètes, dans l'entendement divin qui en connaît à la fois les causes et les lois de production. De même, encore, l'individu est-il capable des développer ces idées inadéquates, en tant qu'elles le caractérisent, à une échelle qui excède négativement les possibilités divines, sous les espèces du délire, de l'erreur, de la superstition, du fantasme, etc., puisque la substance infinie ne saurait avoir la moindre idée inadéquate. Par conséquent, non seulement l'analogie est erronée, mais elle inverse les rapports de puissance: ce que la machine fait de mieux – calculer et combiner des symboles à des échelles et des vitesses inaccessibles à l'homme – elle ne le pense pas; tandis que les pires idées que produit l'homme sont encore pensées, fut-ce inadéquatement ou passivement. Tout au plus peut-on donc, pour l'instant, avancer l'hypothèse que la machine et ses opérations constituent des effets, mais pas des spécifications, de ces modes que sont le corps et l'âme de l'individu humain, au même titre que n'importe quel autre artefact entrelaçant rapports de signes et éléments sensibles (un livre, par ex.). Le problème étant que les effets de ces effets sont peut-être, pour une part, inédits.

1.3. L'analogie dans une perspective freudienne

Dans une perspective freudienne, l'analogie entre l'individualité somatopsychique et l'ordinateur doté d'intelligence artificielle est encore moins envisageable, puisque la psychè est, au départ, constituée par l'ensemble des pulsions que la poussée du somatique contraint l'individu à élaborer comme autant de représentants (*Repräsentanten*) de ses *stimuli* (*Reize*) somatiques.²⁹ Si elle acquiert une relative indépendance à l'égard du corps, c'est, d'une part, parce que ce travail d'élaboration se poursuit pour donner lieu à des représentations (*Vorstellungen*), dans la mesure où l'énergie pulsionnelle s'y investit et où le nourrisson acquiert progressivement la distinction du moi et du non-moi, puis du sujet et de l'objet; d'autre part, parce que ce processus d'élaboration est inséparable de celui de la structuration psychique en systèmes distincts: à partir du Ça (*Es*), une organisation complexe se met en place, conduisant, au fil des années, à la constitution du Moi (*Ich*) et du Surmoi (*Überich*).

²⁹ Cf. Sigmund FREUD, *Pulsions et destin des pulsions*, in *Œuvres complètes. Psychanalyse*, XIII, PUF, Paris 1988, p. 169 ss.

Dire que c'est la poussée du somatique qui contraint l'individu à l'élaboration psychique revient à affirmer que l'énergie dont elle est issue ne relève pas du simple besoin, mais qu'elle est d'emblée porteuse de désir: la pulsion résulte, en effet, de l'érotisation du besoin, ou si l'on préfère, elle exprime et manifeste le fait que chez l'être humain les besoins ne sont pas seulement corporels mais aussi de nature affective.

Or, la nature de la corrélation entre le corps (*Körper*) et l'âme (*Seele*) est pour Freud susceptible, contrairement à ce qu'affirme Spinoza,³⁰ de donner lieu à des interactions: si l'âme se constitue sous la poussée du somatique, son énergie propre peut aussi emprunter la voie inverse de la conversion hystérique ou de la somatisation. Il apparaît donc qu'en l'absence d'une capacité qualitativement transformatrice qui, sous la poussée de ses propres besoins et de sa propre énergie, amènerait l'ordinateur à élaborer des idées représentant animiquement ses «appétits» (ou besoins), une telle machine ne peut ni penser, ni avoir une âme. L'hypothèse que quoi que ce soit en elle relève d'un inconscient et qu'elle puisse être atteinte de maladie mentale est également inenvisageable, pour ne pas dire loufoque. Cela n'exclut cependant pas la possibilité que ses imitations de ces fonctions et caractéristiques humaines ne deviennent, à terme et peut-être rapidement, excellentes. Le «*chatbot*» qui nous a fourni le point de départ de notre analyse semble déjà le faire dans une certaine mesure. Dans ces conditions, comment penser le rapport entre l'affect et son imitation extérieure par le «*chatbot*» ou par les autres machines dites artificiellement «intelligentes»? Peut-on s'attendre à ce que ces imitations affectives artificielles remplissent le même type de fonctions que de véritables affects?

2. L'imitation artificielle des affects

Prendre au sérieux les effets que des machines dotées d'intelligence artificielle peuvent avoir sur leurs utilisateurs implique de saisir que le rapport à la réalité extérieure n'est pas premier dans les mécanismes de la revendication et de la satisfaction pulsionnelle, pas plus qu'il ne l'est dans ceux de l'engendrement et de la reconduction du désir. Pour Spinoza, la finitude du corps, en particulier, et la tendance à persévérer dans son être qui le caractérise comme toute chose finie, sont des conditions amplement suffisantes à la genèse du désir dans l'âme humaine. Pour Freud, ce n'est que de manière seconde, lorsqu'il a déjà mis en place un certain nombre de

³⁰ E, III, 2.

structures facilitant son accès à la rationalité, que l'individu relie la possibilité d'une satisfaction relationnelle à l'existence d'un enjeu réel (relevant d'une réalité externe et non pas simplement interne). De fait, même lorsqu'il est un adulte capable de comprendre le monde qui l'entoure et d'agir efficacement sur lui en se fondant sur les connaissances qu'il en a accumulées, les enjeux réels se doublent toujours pour l'individu humain d'une dimension symbolique et imaginaire, qui trouble la distinction entre réalité mentale et externe. Ainsi, pour comprendre la possibilité qu'un être humain développe une relation affective apparemment dénuée d'enjeu avec un robot conversationnel,³¹ il faut d'abord rappeler ce qu'est une relation affective.

2.1. Nature de la relation affective réelle et artificielle

Pour Spinoza, l'affect est à la fois une variation de la puissance du corps et l'idée confuse de cette affection;³² pour Freud, il est conscient et constitue une décharge (*Abfuhr*) d'énergie psychique. Mettre en relation un affect et un objet n'implique donc nullement, dans les deux cas, que l'objet soit nécessairement la cause de cet affect, ni que l'individu éprouvant cet affect ait une juste représentation de sa cause ou de la relation qui lie l'affect à l'objet. La cause de l'affect peut même être, en réalité, essentiellement interne, bien qu'imaginativement identifiée à l'objet par l'individu. Ainsi, selon la logique affective développée par la génétique des passions d'E, III, il suffit que l'individu imagine l'amour ou la haine d'un objet à son encontre, que cet objet soit humain ou non, pour éprouver en retour, par le mécanisme de l'imitation affective, ce même affect à quelque degré pour l'objet. Dans une perspective freudienne au sens large, les mécanismes du transfert³³ ou de l'identification projective³⁴ relèvent, de la même façon, d'une logique interne beaucoup plus qu'externe (du moins dans un premier temps, au moment de la genèse de l'affect).

On peut donc se demander jusqu'à quel point l'individu est capable de ne pas réagir affectivement à quelque degré lorsqu'il est confronté à la simulation de certains affects par un «*chatbot*» censé jouer le rôle de «double virtuel» ou de «meilleur ami». Lui suffit-il pour cela de savoir qu'il n'a affaire qu'à un programme, que celui-ci collecte

³¹ Cf. Serge TISSERON, *Petit traité de cyber-psychologie. Pour ne pas prendre les robots pour des messies et l'IA pour une lanterne*, Le Pommier, Paris 2018, pp. 163-177.

³² E, III, Définition générale des affects.

³³ Cf. Monique SCHNEIDER, *La cause amoureuse. Freud, Spinoza, Racine*, Seuil, Paris 2008.

³⁴ Mise au jour par M. Klein.

des données à son sujet, est connecté en permanence à l'entreprise qui le commercialise, etc.? Si l'illusion produite par le «*chatbot*» est suffisamment forte, elle risque de produire des affects plus forts que ceux qu'implique l'ensemble de ces savoirs, telles la réserve et la prudence. Les témoignages produits dans l'article du *Monde* cité plus haut laissent à penser que ses utilisateurs les oublient assez vite, voire n'ont pas bien compris que le robot n'a rien d'une entité close sur elle-même et susceptible de n'être connectée qu'avec eux-mêmes ou des amis lors des «conversations» (ou échanges de signes linguistiques) qu'ils ont avec lui.

Dans une relation affective entre individus humains, des idées de variations de puissance corporelle sont rapportées à un objet extérieur par la médiation d'un certain nombre de significations. Cela implique que l'affect a, d'un point de vue mental, une double dimension: d'une part, l'affect a une dimension animique, puisqu'il est pensé par la partie de l'âme³⁵ qui se rapporte au corps et qui contient idées imaginaires et souvenirs; d'autre part, il est associé à des idées de mots qui vont permettre à la fois son expression et son interprétation, et s'inscrit donc dans une langue déterminée parlée par l'individu, qui a été déposée en lui sous la forme d'un ensemble de traces (*vestigia*) sédimentées et remaniées par ses usages successifs dans la durée. L'affect peut donc être communiqué et partagé par différents individus: chacun, certes, l'éprouvera à sa manière singulière et à un degré qui lui sera propre, mais une certaine communauté affective émergera et se constituera, d'une part à partir de la communication linguistique de l'affect, d'autre part, à partir des lois universelles régissant les mécanismes émotionnels de la nature humaine.

Dans l'imitation artificielle d'un rapport affectif humain, cette dimension symbolique et imaginaire n'existe évidemment pas, pas plus qu'une quelconque ébauche de communauté affective. Le *chatbot* organise des ensembles de symboles à partir des règles qui structurent à l'avance le champ d'indétermination qu'il est programmé à affronter et à explorer de manière à le maîtriser progressivement de plus en plus. Le lexique et les usages linguistiques qu'il met en œuvre sont dépourvus de toute épaisseur mnésique ou mémorielle, ils sont empruntés (au système constitué par la langue et ses déclinaisons par un individu humain dans certains documents) et copiés, non appropriés. Les affects apparents qu'il manifeste sont, à ce titre, simulés³⁶ et non à

³⁵ Généralement nommée *animus* (lorsqu'il s'agit de l'homme), et moins souvent *anima* (lorsqu'il s'agit des autres espèces) par Spinoza.

³⁶ Leur extériorité nous paraît justifier le terme de «simulation».

proprement parler imités, car le mimétisme affectif qui règne entre les hommes implique qu'ils éprouvent l'affect imité et l'intègrent ainsi à leur vie animique. C'est d'ailleurs ce mécanisme universel³⁷ qui explique que dans l'esprit de l'utilisateur, qui retire sans doute quelque agrément de l'usage du *chatbot* – détente, illusion de présence amicale ou de personne disparue, gratification affective, soutien ou réassurance narcissique, etc. – la simulation affective risque d'être prise pour une imitation, avec ce que cela suppose d'illusion de communauté affective.

2.2. Aspects affectifs fonctionnels du *chatbot*

Deux aspects du *chatbot* Replika, tel qu'il est décrit dans l'article déjà cité, retiennent en particulier notre attention. Le premier est qu'il est censé garantir un contact «plus facile», plus gratifiant, «moins risqué»³⁸ qu'avec un individu humain.³⁹ Il s'agit manifestement pour son usager d'accéder à un type de gratification narcissique qui s'apparente à un ersatz de reconnaissance, plus simple à obtenir de la part d'un robot que d'un être humain réel, qu'il faut ménager et reconnaître en retour en tenant compte des particularités de sa situation et de son *ingenium*. La fonction affective du robot pourrait être décrite en termes de réassurance ou de soutien narcissiques: telle une béquille affective, il est là pour étayer de façon ludique l'estime et l'amour de soi de celui qui l'utilise. À supposer que cette fonction soit effective, même à un moindre degré qu'avec un autre être humain, cela signifie alors qu'importe plus l'illusion, peut-être pour un temps momentanée et fugitif seulement, que la réalité, qui se trouve escamotée, en même temps que les risques de déconvenues et de blessures narcissiques impliqués par une relation affective interhumaine. Cette dimension narcissique des fonctions affectives de Replika est particulièrement manifeste dans l'idée qu'il deviendrait le «double virtuel» de l'individu: en somme, il s'agit pour lui, en partie, de contempler une image plaisante de soi, doublée du réconfort narcissique qu'il s'administre en réalité lui-même par le truchement d'un logiciel.

³⁷ E, III, 21.

³⁸ EUDES, *Replika*, p. 16.

³⁹ Cette présentation est tendancieuse, le terme de «relation» impliquant, entre individus humains, le rapport à une altérité. Or, celle-ci est ici essentiellement imaginaire, ou réduite aux effets physiques des algorithmes réglant l'acquisition progressive, par la machine, d'une aptitude mimétique mécanique.

À cet égard, Replika fait penser, d'un point de vue animique ou psychique, à un «objet transitionnel» pour adulte, l'équivalent sophistiqué du «doudou» utilisé par les jeunes enfants, et, d'un point de vue social, à une sorte de majordome attentionné supposé veiller au bien-être affectif de son «maître». À ceci près, justement remarqué par S. Tisseron, que contrairement à l'objet transitionnel, il ne permet pas «une élaboration progressive des projections de la personne [...], de telle façon qu'il devienne ensuite inutile». ⁴⁰ Aussi le qualifie-t-il plus volontiers d'«absorbeur d'anxiété» ou de «support d'accordage affectif». ⁴¹ Provisoirement replié sur lui-même et protégé (ou consolé) du réel, son usager connaîtrait le confort d'un «cocon» ou d'un maternage affectif et psychique. ⁴² Il y aurait donc ici l'inverse du phénomène d'«excorporation psychique» décrit par S. Tisseron ⁴³ à propos de l'utilisation d'avatars dans des jeux vidéos en ligne: tandis que dans ceux-ci, l'individu s'identifie à un avatar souvent très différent de lui-même, qui lui permet de vivre dans un monde virtuel des expériences inédites et irréalisables dans la vie réelle, Replika permettrait de réaliser une empreinte psychique partielle de soi sous une forme narcissiquement flatteuse, un «dédoublement» qui serait à l'âme ce que le «*selfie*» retouché est au corps.

Le deuxième aspect remarquable de cette application pour *smartphone* réside dans les possibilités de délégation et de représentation affectives qu'elle semble offrir vis-à-vis des proches de l'individu: à ce titre, Replika ne tient pas seulement de l'étai, mais de l'appendice narcissique. La tâche de cultiver ses relations par des contacts réguliers pourrait ainsi être déléguée au *chatbot*, comme une chose à faire parmi d'autres. Paraît assez significatif le fait que les «conversations» ainsi déléguées ne soient pas enregistrées, et nous ne pouvons faire que des suppositions à ce sujet. Sans doute y a-t-il là des motifs juridiques, la collecte de données personnelles ne pouvant se faire sans l'accord préalable de l'utilisateur du logiciel, accord que n'ont pas donné ses proches. Reste qu'une telle absence d'enregistrement ne peut que produire une illusion accrue d'autonomie du *chatbot* comme de sa prétendue réalité de double virtuel, et qu'elle rend, de fait, le contrôle impossible. Sans doute constitue-t-elle également un facteur

⁴⁰ TISSERON, *Petit traité de cyber-psychologie*, p. 224.

⁴¹ *Ibid.*

⁴² Qualifié de «bain amniotique» par S. Tisseron, *ibid.*

⁴³ Soit de «décorporation du corps propre» et d'«encorporation dans un autre corps», cf. *ivi*, pp. 164-166.

ludique supplémentaire, sous la forme d'une intégration de la surprise et de l'inconnue comme éléments du jeu.

Dans ces deux aspects principaux, Replika réalise partiellement ou esquisse quelque chose comme une double adjonction: adjonction d'un étoi et d'un appendice narcissiques à l'individu, et adjonction de son double virtuel partiel au *cloud*. En un tel programme se projette clairement le fantasme d'une existence démultipliée et immortelle, puisque le double virtuel, contrairement à l'individu humain, sauf opération de suppression expresse, n'est pas censé disparaître. Sous des apparences relativement novatrices et ludiques, une telle invention révèle-t-elle quelque chose des enjeux liés au développement économique de l'I.A.? Sans prétendre les cerner tous ni les approfondir, peut-on du moins en esquisser quelques-uns?

3. Enjeux et conséquences possibles de ce type d'I.A.

L'enjeu le plus manifeste de la mise sur le marché d'un tel logiciel est évidemment économique: Replika, et l'ensemble des logiciels de même type, constituent une tentative nouvelle d'exploitation des attentes affectives et de l'isolement relatif (volontaire ou subi) des individus, par la vente d'un service robotisé d'étayage et de représentation narcissiques. En collectant des données sur ses utilisateurs, le robot permet d'affiner à la fois le «profilage» des consommateurs et l'offre des services susceptibles d'alimenter et de renouveler la consommation. Comme tant d'applications et de divertissements proposés en ligne, il dissimule, en tant que «produit», un marketing qui, comparé aux techniques et aux outils de commercialisation déployés hors internet, paraît fortement intrusif. Or, l'accès aux données stockées dans le *cloud* pourrait, à plus ou moins long terme, ne pas être directement ou seulement l'apanage de l'entreprise qui le commercialise, mais se faire aussi *via* d'autres programmes automatiques dotés d'I.A.: c'est du moins ce qu'implique la notion d'«apprentissage profond» ou de *deep learning*.

Par ailleurs, comme doublure virtuelle partielle du réel adjointe (et peut-être, à terme, intégrée) aux relations humaines, des logiciels du type de Replika pourraient, pour certains individus, contribuer à troubler la frontière du vrai et de l'illusoire. En donnant à l'individu la possibilité d'une délégation de segments d'activité relationnelle qui peuvent se poursuivre sans lui et indépendamment de son contrôle possible, le *chatbot* tend, de fait, à estomper la distinction entre soi et non soi, réel et virtuel, vrai

et faux. Le risque serait alors qu'à force de se plaire dans ces pseudo-échanges virtuels, l'utilisateur en vienne à négliger d'en avoir de véritables, et en perde jusqu'au goût. Les services d'étayage ou de réassurance narcissiques achetés pourraient alors de se payer bien cher, d'autant que l'individu ne sera qu'apparemment valorisé, et que l'attrait réel de sa compagnie n'en sera pas modifié aux yeux des autres. En outre, s'il peut être remplacé, même ponctuellement, par un *chatbot*, le «prix» de son contact relationnel se trouve, de fait, en quelque sorte dévalué ou revu à la baisse, ce qui semble constituer plutôt un facteur d'affaiblissement narcissique.

De fait, si des aspects de nos interactions relationnelles peuvent être délégués, comme s'il s'agissait d'un travail (entretien et développement de son réseau social et de sa visibilité sociale, autopromotion, diffusion d'annonces et de nouvelles, etc.) ou de tâches d'importance secondaire, alors le «moi», sous l'espèce de sa réplique artificielle, est supposé pouvoir se manifester indépendamment du corps propre, mais aussi de l'esprit de l'individu; artificiellement adjoint ou intégré à des relations dont il était auparavant exclu ou absent, il pourrait aussi, vis-à-vis d'autrui, remplir certaines des prestations affectives auxquelles il l'a habitué. Ce n'est, en soi, pas nouveau: une lettre ou un message vidéo remplit aussi, quoique différemment, ce genre d'offices. Mais se trouve manifestement amorcée ou tentée quelque chose qu'il faut bien qualifier d'industrie robotisée des services affectifs, dont les prémises relativement concluantes se dessinent dans l'utilisation de robots conversationnels dans les soins dédiés aux patients atteints de la maladie d'Alzheimer,⁴⁴ ou comme outils thérapeutiques dans certaines prises en charge d'adolescents.⁴⁵

«Développer une relation émotionnelle avec une machine»,⁴⁶ comme invitent à le faire les inventeurs de Replika, est sans doute possible,⁴⁷ au sens où son utilisateur peut être amené à développer un certain nombre d'affects envers elle. Mais, tandis qu'avec une machine non robotisée le développement de ces affects n'est issu que de l'imagination et de la capacité projective de l'individu, les logiciels d'assistance mentale

⁴⁴ *Ivi*, p. 36.

⁴⁵ *Ivi*, p. 271.

⁴⁶ EUDES, *Replika*, p. 16.

⁴⁷ Au point que des démineurs militaires peuvent mettre leur vie en danger pour sauver leur robot de déminage: cf. Serge TISSERON, *An Assessment of combatant Empathy for Robots with a View to avoiding inappropriate Conduct in Combat*, in Ronan DOARE, Didier DANET, Jean-Paul HANON et Gérard de BOISBOSEL (dir.), *Robot on the Battlefield, Contemporary Issues and Implications for the Future*, Combat Studies Institute Press, Forth Leavenworth, Kansas 2014, pp. 165-180.

ou affective dotés d'I.A. sont programmés pour l'induire et renforcer ce mécanisme projectif, si ce n'est engendrer une dépendance, sous l'espèce de besoins nouveaux. Aussi, excepté peut-être dans certaines situations de soins, un tel type de relation paraît peu souhaitable. Les risques de confusion, de perte de contact avec la réalité ou de perte du sens des relations humaines paraissent, en effet, élevés par rapport aux bénéfices apportés par l'usage de ce type de robots. Les inventeurs de *chatbots* et autres agents artificiels d'assistance relationnelle et mentale vont pourtant tout faire pour y inciter les individus, ne serait-ce que parce que cela leur permet et de collecter une masse de données phénoménale et de développer leur activité économique. Ces données, convenablement traitées et analysées, leur donnent la possibilité, par la médiation de ces logiciels, d'apprendre à connaître les affects et les mécanismes émotionnels de leurs utilisateurs.⁴⁸ Concrètement, ils pourront alors chercher à interagir à distance avec eux, et viser à influencer, voire à «contrôler» leurs affects. Les robots dotés d'I.A. pourraient ainsi inaugurer de nouvelles modalités de captation affective.

Dans la mesure où les affects déterminent le plus souvent ce que l'individu pense, c'est peut-être en réalité une ambition sociale et politique de détermination des opinions individuelles et finalement, de modelage de ce que É. Balibar appelle le «transindividuel»⁴⁹ qui pourrait ainsi avoir pris, avec les développements de l'I.A., une forme inédite. En effet, si le transindividuel est constitué par l'ensemble des pratiques qui permettent à une multitude de constituer un corps tel qu'il soit un quasi Individu d'individus, les modifications susceptibles d'être apportées par la banalisation et la massification de l'utilisation de robots dotés d'I.A. pourraient être profondes. Or, les entreprises regroupées sous l'appellation «GAFAM» n'ont pour l'instant pas d'équivalent en Europe: seule la Chine, actuellement, paraît en mesure de rivaliser avec elles. Ce sont donc, comme l'ont montré les révélations de l'affaire Snowden, non seulement des formes nouvelles⁵⁰ d'espionnage et de guerre industrielle qui se mettent en place dans les relations interétatiques, mais peut-être aussi, à termes, de nouvelles modalités de sous-traitance économique et de domination politique.

⁴⁸ Ainsi, l'étude des filtres utilisés sur Instagram permettrait de prédire les dépressions nerveuses: cf. Andrew G. REECE et Christopher M. DANFORTH, *Instagram photos reveal predictive markers of depression*, in «EPJ Data Science», 2017, volume 6, article 15. Scruter les différents réseaux sociaux utilisés par la personne est déjà une étape avérée du recrutement professionnel dans bon nombre d'entreprises.

⁴⁹ Étienne BALIBAR, *Spinoza politique. Le transindividuel*, PUF, Paris 2018.

⁵⁰ Par leurs supports technologiques et leurs méthodes, non par leurs visées ou leurs contenus.

Enfin, les conséquences pourraient en être aussi, et peut-être surtout, psychiques et morales: les individus sont incités à une narcissisation qui, paradoxalement, pourrait bien être fragilisante, et favoriser la substitution de relations de services et d'échanges de prestations aux relations affectives véritables entre individus humains. S'ils se mettent à déléguer de plus en plus à des robots des tâches apparemment subalternes de représentation affective ou sociale, les individus vont être habitués à penser différemment leur propre moi, voire à définir tout autrement leur identité.⁵¹ La personnalité ou l'*ingenium* d'un individu peut-il être considéré comme immanent à l'ensemble des relations qu'il initie, même s'il ne les vit pas et ne les perpétue pas lui-même? Ce déplacement de l'importance sociale vers la surface ou l'étendue de la sphère sociale, qui repose aujourd'hui sur des critères quantitatifs bien plus que qualitatifs (nombre de «like», de «followers» et d'«amis», etc.) va dans le sens d'un élargissement de la représentation du «moi» (par l'introduction de nouvelles modalités représentatives et relationnelles), et d'une perte au moins partielle de sa profondeur et de sa vérité. Le redoublement de la surface sociale par une surface virtuelle pourrait bien être ainsi un facteur d'accentuation de la séparation de l'être et du paraître par lequel se manifeste toute culture. Exister pourrait-il un jour alors signifier multiplier les surfaces d'exposition, déléguer les expressions de soi, coordonner de loin sa vie sociale? C'est là un danger, celui de la superficialité et de la simulation, dont on a pu, dans le passé, accuser également l'écrit, la photographie, la vidéo. Mais, ce qui est inédit, c'est le caractère apparemment indépendant, autonome, de ce dédoublement-redoublement, alors même qu'il est, pour l'instant,⁵² constamment surveillé, analysé, scruté par ceux qui l'ont initié.

Le narcissisme et les instruments de sa satisfaction ne sont pas choses nouvelles: miroirs, portraits, mises en scène de soi, etc., sont intrinsèques à toute culture. Aucun de ces instruments, cependant, n'était apparu jusqu'ici qui puisse paraître constituer un substitut ou une menace pour les relations humaines réelles, sur lesquelles il était toujours possible de compter pour ramener l'individu à la raison en cas de délire trop prononcé. Seul le mythe avait pu rêver d'instruments anthropomorphes de représentation des individus: mais le miroir parlant ou réflecteur de la vie d'autrui du

⁵¹ Cf. TISSERON, *Petit traité de cyber-psychologie*, pp. 173-177.

⁵² Le «*deep learning*» serait, à terme, susceptible de rendre cette autonomie plus véritable: cf. MALABOU, *Les métamorphoses de l'intelligence*. Nous ne nous prononcerons pas sur cette possibilité.

conte⁵³ semble, pour partie, devenir réalité. De même, les outils et techniques de captation des affects à des fins économiques et politiques n'ont rien de nouveau. Mais si, comme l'a montré Freud, les ressorts de l'identification des masses à un chef politique sont de nature narcissique,⁵⁴ alors l'illusion de communauté affective que peuvent propager certains robots dotés d'I.A. pourrait s'avérer une arme de propagande assez retorse, en renforçant à demeure, et de façon quasi permanente, les mécanismes d'une telle identification. Ainsi, à l'ambition de la maîtrise de l'âme et aux tentatives d'une domination des corps, succèderait le rêve d'un modelage affectif contrôlé du transindividuel.

Bibliographie

Étienne BALIBAR, *Spinoza politique. Le transindividuel*, PUF, Paris 2018.

Yves EUDES, *Replika. L'application mobile qui fait parler les morts*, in «Le Monde», 29 juillet 2017, p. 16.

Sigmund FREUD, *Pulsions et destin des pulsions*, in *Œuvres complètes. Psychanalyse*, XIII, PUF, Paris 1988

Sigmund FREUD, *Psychologie des masses et analyse du Moi*, in *Œuvres complètes. Psychanalyse*, XVI, Paris, PUF, 1991.

Jean-Gabriel GANASCIA, *La querelle de l'âme des machines*, in Jean-Luc GUICHET (dir.), *De l'animal-machine à l'âme des machines. Querelles biomécaniques de l'âme (XVIIe-XXIe siècle)*, Publications de la Sorbonne, Paris 2010, pp. 169-183.

Jean-Gabriel GANASCIA, *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle?*, Seuil, Paris 2017.

Jean-Gabriel GANASCIA, *Intelligence artificielle: vers une domination programmée?*, Le cavalier bleu, Paris 2017.

⁵³ Cf. *Blanche-neige, La belle et la bête*.

⁵⁴ Sigmund FREUD, *Psychologie des masses et analyse du Moi*, in *Œuvres complètes. Psychanalyse*, XVI, Paris, PUF, 1991.

- Hubert GUILLAUD, *À qui les interfaces vocales s'adaptent-elles?*, «Le Monde», 20 janvier 2019, <https://www.lemonde.fr/blog/internetactu/2019/01/20/a-qui-les-interfaces-vocales-sadaptent-elles/>.
- Ray KURZWEIL, *La singularité est proche. Quand les humains transcendent la biologie*, Viking Press, New York 2005.
- Catherine MALABOU, *Les métamorphoses de l'intelligence. Que faire de leur cerveau bleu?*, PUF, Paris 2017.
- Andrew G. REECE et Christopher M. DANFORTH, *Instagram photos reveal predictive markers of depression*, in «EPJ Data Science», 2017, volume 6, article 15, <https://epjdata.science.springeropen.com/articles/10.1140/epjds/s13688-017-0110-z#citeas>.
- Monique SCHNEIDER, *La cause amoureuse. Freud, Spinoza, Racine*, Paris, Seuil, 2008.
- Baruch SPINOZA, *Éthique, Œuvres III*, Garnier-Flammarion, Paris 1965.
- Serge TISSERON, *Petit traité de cyber-psychologie. Pour ne pas prendre les robots pour des messies et l'IA pour une lanterne*, Le Pommier, Paris 2018.
- Serge TISSERON, *An Assessment of combatant Empathy for Robots with a View to avoiding inappropriate Conduct in Combat*, in Ronan DOARE, Didier DANET, Jean-Paul HANON et Gérard de BOISBOSEL (dir.), *Robot on the Battlefield, Contemporary Issues and Implications for the Future*, Combat Studies Institute Press, Fort Leavenworth, Kansas 2014, pp. 165-180.
- Cédric VILLANI, *Donner un sens à l'Intelligence Artificielle. Pour une stratégie nationale et européenne. Mission parlementaire du 8 septembre 2017 au 8 mars 2018*, <https://cnnnumerique.fr/publication-rapport-Villani-IA>.
- Christian WOLFF, *Gesammelte kleine philosophische Schriften*, G. Olms, Hildesheim, New York 1981.