



HAL
open science

Du gène à l'octet : la communication phygitale pour une utilisation responsable de l'Intelligence Artificielle dans le domaine médical

Christophe Denis, Judith Nicogossian

► To cite this version:

Christophe Denis, Judith Nicogossian. Du gène à l'octet : la communication phygitale pour une utilisation responsable de l'Intelligence Artificielle dans le domaine médical. 2019. hal-02313561

HAL Id: hal-02313561

<https://hal.sorbonne-universite.fr/hal-02313561v1>

Preprint submitted on 11 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Du gène à l'octet : la communication phygitale pour une utilisation responsable de l'Intelligence Artificielle dans le domaine médical.

DENIS Christophe, NICOGOSSIAN Judith

¹Maitre de conférences, Laboratoire d'Informatique LIP6, Sorbonne Université, christophe.denis@lip6.fr

²Anthropobiologiste, Sociolinguiste

Résumé :

Depuis 2010, l'Intelligence Artificielle (IA) connexionniste fondée sur de l'apprentissage machine produit des résultats impressionnants, principalement dans les domaines de la reconnaissance de forme, du traitement naturel du langage et de la perception, succédant à la domination de l'IA symbolique centrée sur le raisonnement logique et formel. L'IA bouleverse des pans entiers de la société humaine et du vivant. Elle est un rouage de la mutation en cours de la médecine. Ces systèmes dits intelligents peuvent produire un certain nombre d'effets néfastes chez les usagers de la santé. Véritables colosses au pied d'argile, ils possèdent des failles (e. g. les biais afférents aux données dont elles se nourrissent) et peuvent également renforcer les inégalités en santé. Alors, au contraire d'augmenter le pouvoir d'agir des usagers de la santé sur leur parcours de soin, le danger serait de réduire le pouvoir d'agir au détriment des usagers. Plutôt que tenter d'adapter l'outil dans ses contraintes et ses limitations sur les modes de communication existants à ce jour entre les humains, nous proposons de définir un mode de communication phygitale spécifique entre l'homme et la machine.

I. Introduction

La mise au point d'une machine électronique dotée d'une capacité de raisonnement semblable à celle des être humains a motivé la naissance durant l'été 1956 de l'Intelligence Artificielle (IA) lors de la devenue célèbre conférence de Dartmouth. En particulier, la cybernétique s'est définie comme la discipline étudiant et mimant le raisonnement humain à l'aide de réseau de neurones artificiels. La relative simplicité des réseaux de neurones, par exemple le perceptron, était en profond décalage par rapport aux fortes attentes initiées par cette jeune discipline. Cette grande désillusion a provoqué une forte réduction des financements publics et privés pour ce champ de l'IA au profit de l'approche symbolique, basée sur le raisonnement formel et la logique à travers notamment des systèmes experts. Le fort intérêt industriel pour ces systèmes experts exprimé au début des années 1980 a lui aussi fortement diminué dix années plus tard en raison des limites de cette approche.

On assiste depuis 2010 à un renouveau de l'Intelligence Artificielle porté par les résultats spectaculaires de l'apprentissage machine dans le champ de la reconnaissance de formes et de la prédiction statistique. Le principe de ces méthodes est de trouver des combinaisons entre un nombre potentiellement de données à l'aide de corrélations statistiques pour fournir une prédiction ou un diagnostic.

Les enjeux de l'IA s'inscrivent aujourd'hui dans une mutation radicale de nos sociétés, qui doivent faire à face à des problèmes de gouvernance (e.g. démographie, accès au soin) et d'écologie en lien (e.g. amenuisement des ressources). Cette imbrication forte entre IA et nos activités humaines obligent à poser la question de la responsabilité numérique. Cet article aborde la problématique de l'utilisation responsable de l'IA dans le domaine de la santé.

II. Intelligence Artificielle et Santé

Que l'on souhaite ou le déplore, l'Intelligence Artificielle semble être devenu le nouvel eldorado pour trouver de nouveaux relais de croissance sur une planète dont les ressources naturelles sont menacées de *burn-out* puisqu'épuisées par une démographie humaine toujours croissante (11 milliards d'humains en 2050). Cette évolution de la démographie s'accompagne d'un (i) vieillissement des populations occidentales et d'une (ii) augmentation de l'espérance de vie en bonne santé (iii) d'une chronicisation des maladies. Ces trois enjeux placent frontalement les systèmes de santé face à la nécessité d'élaborer des mesures concrètes. Le paradigme de la médecine des 4P semble être le moyen pour gérer le compromis entre la réduction des coûts et l'amélioration de la qualité des soins. Cette promesse s'articule autour de quatre piliers :

- *préventif* : informer, éduquer des bonnes conduites à tenir (e.g. alimentation) ;
- *prédictif* : caractérisation pré-symptomatique du risque de développer telle ou telle maladie sur des facteurs génétiques ;
- *personnalisé* : traitement de soin adapté à la culture et/ou au métabolisme de l'individu ;
- *participatif* : contribution active du patient (et de ses aidants) dans son parcours de soin.

L'Intelligence artificielle est la clé de voute technologique pour la mise en place d'au moins une partie de cette médecine 4P, qui implique l'élaboration de nouveaux droits et devoirs numériques. Avant de détailler dans la prochaine section les effets potentiellement néfastes de l'IA, sur lesquels nous devons exercer une vigilance éthique, il est utile de rappeler que cette volonté de développer des systèmes automatisés dans le domaine de la santé, en interaction avec les utilisateurs, n'est pas neuve. Rappelons, dans le contexte médical :

- *l'agent conversationnel Eliza* :

Le programme Eliza a été développé par Joseph Weizenbaum chercheur en Informatique au MIT¹. Ce programme simule une conversation avec un psychologue rogerien. Il s'agit d'une branche de la psychothérapie orientée sur la personne, basée sur une relation empathique entre le thérapeute (ici Eliza) et le patient. Le principe de l'agent conversationnel est simple. Il reformule sous forme de questions les affirmations du patient. Par exemple, l'agent pose la question « *Can you explain what made you unhappy* » lorsque le patient affirme auparavant. « *It's true. I am unhappy* ». L'agent répond « *I understand* » lorsque celui-ci ne peut plus reformuler de questions adaptées à l'affirmation du patient.

Certains utilisateurs ont pensé avoir à une psychologue, le programme Eliza ayant ainsi passé dans l'absolu le test de Turing². Mais également, d'autres usagers ont développé une dépendance à l'agent conversationnel (« effet Eliza »).

Les raisons pour lesquelles les patients portent plaintes contre les médecins ne dépendent pas d'une erreur de diagnostic, mais avant tout d'un sentiment de dégradation de la qualité de la communication thérapeutique, verbale et non-verbale. D'autres études défendent une corrélation forte entre la bienveillance du médecin et l'efficacité de la guérison. En santé, la relation empathique entre le médecin et le

¹ J. Weizenbaum : « ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine », Communications of the Association for Computing Machinery, 1966.

² A. Turing, « Computing Machinery and Intelligence », Mind, 59, 236, pp. 433-460, 1950.

patient joue un rôle important, même si le processus ne possède encore de preuves, dans la guérison du malade, à l'extérieur du terme générique de « placebo ». Ces analyses montrent le besoin crucial de définir et de spécifier la communication entre les humains et l'IA pour éviter les travers du passage de la médecine à l'acte sur des supports technologiques.

- *le programme d'identification de bactéries MyCin :*

Le système expert MyCin³, développé dans les années 70 à l'Université de Stanford, a pour fonction d'identifier les bactéries causant des infections graves, telles que la bactériémie et la méningite, et de recommander des antibiotiques, la posologie étant adaptée au poids du patient.

MyCin n'a jamais été utilisé dans la pratique, non pas pour des performances faibles, mais principalement pour des problèmes liés à l'utilisation des ordinateurs en médecine, juridique et éthique.

Les limites de ces utilisations sont en lien au diagnostic posé par la machine et au remplacement du médecin par la machine. Ces problèmes sont encore d'actualité. Les enjeux sont importants. Que deviendrait un monde où l'on préfère avoir affaire à un médecin IA plutôt qu'un médecin humain ? Que penser de l'effet addictif d'Eliza, cette machine très sommaire décrite plus haut ? Comment extrapoler cet effet Eliza sur une machine plus perfectionnée ? Rappelons que le smartphone, par exemple, ne pose pas la question de l'acceptabilité, mais celle de l'addiction.

III. Le talon d'Achille de l'Intelligence Artificielle pour la santé

Les méthodes d'apprentissage machine permettent d'obtenir des prédictions de bonne qualité conduisant à des prises de décisions. Il ne faut pas perdre de vue la critique d'un « exceptionnalisme médical »⁴, là où les systèmes se comportent à basse prévalence alors qu'ils sont présentés à haute prévalence (e.g. les systèmes experts d'imagerie médicale). Les méthodes d'apprentissage utilisent des données historiques inévitablement entachées de biais, qui peuvent négliger des cas exceptionnels. La communication entre le tiers technologique et le médecin est donc indispensable pour contrôler ces résultats. Des questions juridiques concernent le partage de responsabilité entre le médecin et le tiers technologique, notamment celle de la possibilité de remise en cause du diagnostic automatisé, traitant des dilemmes éthiques ; du droit à l'erreur ; et de la question juridique du statut des machines (e.g. droit moral).

Sans négliger la gestion des biais cognitifs, inhérents à tout dispositif explicatif, d'indéniables problèmes de confiance et d'éthique par rapport à l'utilisation de cet outil sont également posés. Le talon d'Achille technique de ces méthodes est qu'il est difficile d'expliquer avec un raisonnement adapté au niveau de connaissance d'un patient ou d'un médecin les raisons conduisant à la prise de décision. La prédiction est la résultante d'une agglomération statistique d'un nombre élevé de signaux faibles dont l'humain peine à donner du sens et qui peut être en contradiction avec l'expertise médicale⁵.

³ B. G. Buchana, E.H Shortliffe (eds.): « Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. », 1984.

⁴ T. Lancet, « Is digital medicine different » Volume 392, Issue 10142, 95, 2018.

⁵ C. Denis, J. Nicogossian, « Explicabilité et convivialité d'outils de diagnostic médical basé sur de l'IA », Journées plénières du GdR-IA, Groupe de travail Explicabilité, Orléans, France, 2019,

De nombreux travaux en Intelligence Artificielle, principalement en informatique, portent sur l'explicabilité des méthodes d'apprentissage machine^{6 7 8} : *Pour autant de quelle explication parle-t-on ? Quel est l'impact de cette explication sur le destinataire ?* Il convient tout d'abord de préciser ce que l'on cherche à expliquer et le destinataire de l'explication, par exemple :

- le fonctionnement de la méthode d'apprentissage en elle-même, (e.g. celui du réseau de neurones profond) destinée à améliorer la connaissance scientifique en mathématiques et en informatique ;
- le sens et les propriétés des résultats produits par la méthode d'apprentissage, destinés aux parties prenantes de ces systèmes.

Sur le plan technique, l'explication de l'explicabilité demeure un véritable casse-tête, et ce, même en décidant d'intégrer à la conception un programme éthique (*ethics by design*). Les décisions du système de *machine-learning* (ML) ne sont pas traçables. Cette volonté semble par ailleurs utopique, quel mécanisme de décision pourrait rendre compte d'une explicabilité parfaite ? Certainement pas la pensée humaine, qui, si elle se nourrit de règles et de lois pour structurer ses échanges en société, agit également en respect de ses instincts. Le droit pénal entend une légitimité de la réaction imprévisible – e.g. l'impondérable face à un danger – avec la notion de préméditation, ou encore celle du bénéfice du doute. Rappelons ici la loi de l'économiste Goodhart⁹ qui indique que lorsqu'une mesure devient un objectif, elle cesse d'être une bonne mesure.

IV. Communication phygitale entre le tiers technologique et le vivant

Plutôt que tenter d'adapter l'outil dans ses contraintes et ses limitations sur les modes de communication existants à ce jour entre les humains, nous proposons de définir un mode de communication phygitale spécifique entre l'homme et la machine. Notre objectif de recherche consiste (i) à définir la qualité de la relation de communication singulière entre le tiers technologique et le vivant, sur le plan de son adaptabilité ; puis (ii) à proposer un protocole de communication phygitale qui permette de résoudre les paradoxes de l'explicabilité, tels que présentés ci-dessus.

La célèbre école de Palo Alto avait défini une anthropologie de la communication en distinguant deux types de communication¹⁰ :

1. la communication *digitale* (analytique, logique, et précise). Elle explique et interprète : elle utilise les codes verbaux ;
2. la communication *analogique* c'est-à-dire affective, plus floue, utilise des symboles : Cette communication est essentiellement non verbale' comprise de tous.

Ce qui est décrit comme notre entrée dans une nouvelle ère digitale invite en permanence l'Homo Sapiens à modifier son comportement, à rejouer ses émotions, proposant une refonte des usages de communication. Un troisième mode de communication entre les tiers technologique et les humains émerge, qui ne doit plus relever :

⁶ C. Denis, Franck Varenne, « Interprétabilité et explicabilité pour l'apprentissage machine : entre modèles descriptifs, modèles prédictifs et modèles causaux. Une nécessaire clarification épistémologique », CNIA 2019, 2019.

⁷ B. Herman, « The Promise and Peril of Human Evaluation for Model Interpretability », Thirsty-first Conference on Neural Information Processing Systems, 2017.

⁸ T. Miller, « Explanation in artificial intelligence: Insights from the social sciences », Artificial Intelligence, 2019.

⁹ C. Goodhart, « Money, Information and Uncertainty », London, Macmillan Press, 1975

¹⁰ P. Watzlawick, P., et al., « Une logique de la communication », trad. Seuil, Norton, [1967] 1972

- ni d'une interaction entre hommes – où la gestion juridique se base sur le droit humain et plus précisément s'inscrit dans une échelle de valeurs axiomatiques des comportements autorisés/interdits en lien à des instincts réprimés/tolérés/acceptés ;
- ni d'une interaction machine-machine – où les dommages collatéraux minimes/majeurs sont jugés en fonction de leur impact sur le réel du point de vue matériel et/ou des dommages corporels engendrés. Pourtant, derrière la machine, se trouve toujours un humain.

Jusqu'à présent la nature – définie par le champ des possibles – s'oppose à la culture – entendue comme la réduction de champ des possibles. Depuis Descartes, puis Diderot, la place du langage prépondérante se trouve au service de la compréhension de la nature, ainsi traduite par des signes.

Avec l'IA les signes du langage subissent une double opération de mise en numérisation, d'une part une syncrétisation des différentes doctrines et des différentes théories et d'autre part une opération de formatage, œuvrant à la disparition du sens pour le combiner en langage binaire. L'interprétabilité des contenus à l'arrivée doit être garantie, et sembler intuitive, pourtant sur cette recombinaison formatée du sens ; ce dernier par définition devient totalement manipulable ¹¹.

Il est force de constater par la communauté scientifique – neurologues, neuroscientifiques, anthropologues etc. - engagée sur la compréhension du fonctionnement du cerveau et ses mécanismes adaptatifs, qu'une dés-évolution menacerait les utilisateurs des interfaces du numérique, ne favorisant ni la créativité, ni l'action, pour au contraire exploiter des mécanismes d'asservissement, comme ceux de l'addiction. « Plus nous nous livrons comme des machines plus nous nous comportons comme des machines »¹². Si un signe peut difficilement être compris comme un signal, les émotions humaines peuvent difficilement être réduites à un signal.

L'hypertrophie de la part technologique ne doit se faire ni au détriment du langage ni au détriment du sens ; en respect de l'ensemble des mécanismes adaptatifs du vivant et des émotions, qui sur le plan de l'évolution, sont essentielles à l'intelligence humaine. Sans elles, - justes dans le savoir -, nous serions prisonniers d'une incapacité à décider. Les sentiments ont pour fonction de nous permettre d'exercer des choix adaptatifs¹³.

La mise en place de ce mode de communication phygitale nécessite de définir le traitement d'un message numérique à deux « entités » : humain et machine. En plus des niveaux de soi à soi et de soi à l'autre, ce mode de communication phygitale permet des rencontres inédites de soi à la machine, de l'autre à la machine. Il ne doit en oublier aucun.

Pour rappel, la situation linguistique d'un mode de communication comporte (i) le porteur de l'énoncé (énonciateur), (ii) le sujet de l'énoncé, (iii) l'énoncé et (iv) le(s) destinataire(s) du message, c'est-à-dire :

1. (i) et (ii) - dans une situation de communication entre deux protagonistes humains le porteur de l'énoncé n'est pas forcément le sujet de l'énoncé ; dans la communication phygitale le porteur de l'énoncé n'est jamais le sujet de l'énoncé, qui utilise le « je ».

¹¹ B. Bachimont, « Patrimoine et Numérique : technique et politique de la mémoire ». INA, Paris, 2017.

¹² J-M. Besnier, France Culture Documentaire "Wellness", <https://www.franceculture.fr/emissions/creation-air/wellness>, 2018

¹³ A. Damasio, « Descartes' Error: Emotion, Reason, and the Human Brain », 1994.

2. (i) et (ii) - dans la communication phygitale le sujet de l'énoncé, qui utilise le « je » n'est jamais sujet, car la machine ne possède ni finalité ni intentionnalité propres.
3. (iii) et (iv) – dans la situation phygitale les destinataires sont trompés par leur sens : le sujet IA n'est pas un sujet ; l'énoncé, à raison d'une immédiate interprétabilité, a subi les manipulations inhérentes au texte numérique ; ils sont coupés d'une partie de la communication non-verbale¹⁴ ; leurs données sont analysées et exploitées ; leurs instincts sont quantifiés ; enfin, ils se trouvent exposés à une techno-addiction par des subterfuges savamment réfléchis (e.g. *comportemental marketing*).

La question de l'altérité dans la machine radicalise la réflexion de Levinas « *La relation est asymétrique à autrui je ne dois pas attendre de réciprocité* ». D'un point de vue éthique, la distance réduit l'empathie et la responsabilité « *Le moi, devant autrui, est infiniment responsable* »¹⁵. Une illusion de réciprocité est le piège tendu par les concepteurs de la machine, se jouant des mécanismes d'addiction (satisfaction-récompense-frustration).

La spécification du mode de communication « phygitale » n'est donc pas un concept creux, dont nous avons dégagé quelques pistes : il permettra de contrôler l'application IA et de gérer d'éventuels dilemmes éthiques.

Conclusions et perspectives

L'Intelligence Artificielle est le vecteur technologique permettant d'accompagner une mutation technologique, sociétale et au-delà civilisationnelle. L'utilisation de l'IA en santé est la source potentielle d'effets néfastes qu'il s'agit de contrôler : par exemple, perte d'autonomie du patient et du médecin, traitement de cas exceptionnels et rares, responsabilité juridique, la perte de compétence du corps médical. Pour rejoindre la notion de convivialité de l'outil théorisée par Ivan Illich, il répond aussi au besoin d'autonomie du vivant par rapport aux outils qui modèlent nos vies^{16,17}.

Nous avons présenté dans cet article un nouveau de cadre de communication entre l'humain et le tiers technologique. Il est basé sur l'hypothèse suivante : les effets indésirables de l'IA resteront imparfaitement résolus et contrôlés si l'on se focalise uniquement sur les propriétés intrinsèques de l'application IA (par exemple l'interprétabilité et l'explicabilité) sans se soucier de l'impact sur le vivant. Les perspectives en cours de cet article sont d'instancier ce cadre de communication phygitale sur plusieurs types d'applications médicales utilisant de l'IA.

¹⁴ Reste à définir laquelle pourrait être retraduite par la machine.

¹⁵ E. Levinas, « Éthique et Infini », (dialogues d'Emmanuel Levinas et Philippe Nemo), Paris, Fayard, coll. « L'Espace intérieur », 1982

¹⁶ I. Illich, « La Convivialité », Seuil, 1973

¹⁷ I. Illich, « Némésis médicale », Seuil, 1975