

IA et coaching, une réflexion éthique et prospective sur leurs apports réciproques

Michel Cezon & Thierry Ménissier

Revue européenne de coaching, n°10, avril 2020,

accessible à l'URL : <https://revue-europeenne-coaching.com/numeros/numero-10-04-2020/ia-et-coaching-une-reflexion-ethique-et-prospective-sur-leurs-apports-reciproques>

Michel Cezon, ingénieur en Intelligence Artificielle, coach professionnel certifié spécialisé dans l'accompagnement des startups technologiques, fondateur et dirigeant de Cogiteo, professeur associé en entrepreneuriat à Université Grenoble Alpes, membre de la Commission IA et Coaching de l'International Coaching Federation, michel.cezon@cogiteo.net

Thierry Ménissier, Professeur des Universités, Philosophie, Institut de Philosophie de Grenoble (IPhiG), responsable de la chaire « éthique & IA », institut interdisciplinaire d'intelligence artificielle MIAI, Université Grenoble Alpes, thierry.menissier@univ-grenoble-alpes.fr

Résumé :

Si la nature de l'intelligence représente une question ancienne, la notion d'une intelligence artificielle (IA) est plus récente et fascine. Qu'une machine soit effectivement intelligente, c'est-à-dire qu'elle puisse décider par elle-même, puisse créer, rêver, etc., cela fait débat. Pour autant, l'IA tend à coloniser tous les domaines de la connaissance, de l'appréhension de notre environnement, à son interprétation et la prise de décision, à nos actions. Dans cette contribution, dans le contexte d'une recherche de l'éthique appropriée à l'IA et en dépit du caractère apparemment iconoclaste de cette position, nous voulons examiner si et comment cette dernière peut (en dépit de ses limites actuelles) apporter à l'activité de coaching, définie comme un accompagnement ponctuel d'un individu ou d'un groupe pour permettre une prise de conscience, un changement de posture, pour amorcer et/ou mettre en œuvre un changement vers un objectif ciblé et/ou mieux traverser une situation difficile. Et également considérer la réciproque, l'apport possible du coaching à l'IA. Pour ce faire, il convient d'examiner les relations humains-machines, notamment à travers le prisme de la projection émotionnelle à laquelle donnent lieu ces dernières.

IA et coaching, une réflexion éthique et prospective sur leurs apports réciproques

Introduction

Dans cet article, nous poursuivons plusieurs objectifs. Premièrement, de manière informative, nous voulons commencer à dresser un état des lieux des avancées de l'IA en

coaching ; cette première dimension constitue le point de départ des trois autres : l'apparition récente de dispositifs techniques, présentés comme des solutions qui permettraient aux usagers de bénéficier d'un accompagnement personnalisé machiniquement assisté, interroge, déroute et inquiète tout à la fois, tant du côté de l'informatique ou de la robotique que du coaching. En effet la relation coach/coaché est intuitu personae et touche à l'intime quand les avancées de la technologie fascinent de par leur universalité. Deuxièmement, à la lumière de ces éléments, il s'agit de préciser les modalités ainsi que les finalités du coaching, notamment en regard du développement personnel d'une part, et de la philosophie pratique ou appliquée, de l'autre. Troisièmement, de manière prospective, notre contribution vise à souligner quels sont les apports et les limites de l'IA en coaching. Enfin, quatrièmement, nous entreprenons, dans une perspective éthique, d'argumenter sur les opportunités et les dangers possibles de l'assistance machinique pour l'accompagnement existentiel ou la performance professionnelle des personnes.

Pour réaliser ces objectifs, nous nous attacherons d'abord à définir les termes en présence. Nous préciserons ensuite les propositions actuelles d'IA de coaching et les axes de développement possible, en montrant notamment que les dispositifs actuels, qui ne relèvent pas nécessairement de l'intelligence artificielle précisément entendue, fonctionnent comme des « substituts IA au coaching » partiellement efficaces. Enfin, nous évoquerons la déontologie du coaching et la perspective d'une éthique de l'IA, en envisageant les relations entre les deux : la déontologie du coaching peut-elle se traduire dans une « éthique de l'IA » ? Celle-ci concernerait à la fois les concepteurs des programmes, les diffuseurs des applications, et les usagers eux-mêmes ?¹

1. IA et intelligence naturelle, l'état actuel de leurs relations :

La notion d'intelligence désigne la capacité de penser, mais non dans l'extension la plus grande du terme. L'intelligence renvoie en effet à la pensée consciente et efficace, capable de développer une connaissance du moi ou du monde et exprimable sous une forme logique ou émotionnelle. Durant des siècles, on avait dans la tradition de philosophie occidentale pris la double habitude de confondre intelligence et rationalité et de centrer l'intelligence sur l'activité

¹ Cet article est le fruit des recherches effectuées dans le cadre de la chaire « éthique & IA » de l'institut pluridisciplinaire en intelligence artificielle MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

du sujet individuel. Mais depuis les années 1960, sous l'effet du développement de plusieurs disciplines telles que la psychologie, la sociologie et l'éthologie, et en tenant compte de l'apport d'autres traditions (par exemple la philosophie orientale), la définition de l'intelligence s'est « élargie » : elle en vient à désigner les formes d'activité intellectuelle et sensible qui permettent l'adaptation et la compréhension du monde par tous les individus possibles (humains et animaux, bien entendu, mais selon une tendance actuelle également végétaux) et en fonction de leurs relations.

On entend par « artificiel » ce qui est créé par l'humain à partir de l'invention technique, de la création artistique ou de la connaissance rationnelle (scientifique ou philosophique) : un outil ou une machine, une sculpture ou une mélodie, un théorème ou un concept peuvent ainsi être dits « artificiels » bien qu'ils soient engendrés par un être naturel, l'humain, et que, s'ils sont efficaces, pertinents ou performants, ils peuvent apparaître comme bien insérés dans la nature, ou adéquats à elle.

Le concept d'une « intelligence artificielle », enfin, apparaît dans les années 1950 avec des algorithmes qui essayent par le calcul de simuler le comportement humain. Alan Turing (1912-1954) a le premier proposé une expérience de pensée valant comme un test destiné à valider « l'intelligence » d'une IA par un dialogue *via* un écran, sous la forme d'un jeu de parloir qu'il a nommé « jeu d'imitation » (*Imitation game*) [Turing 1950]. Le test imaginé par Turing dans son article pose, avec une certaine profondeur philosophique, des questions de nature variée relativement au fonctionnement de l'intelligence, au langage, à l'interaction entre locuteurs humains [Dascal 2005 ; Le Blanc 2014]. Bien qu'il n'ait pas été conçu dans l'intention de fournir une telle application, depuis 1990 le test donne lieu à un concours (le Prix Loebner, doté de plusieurs milliers de dollars) qui met aux prises des agents conversationnels ou « dialogueurs » (*chatbot*) satisfaisant ces critères : si un tiers environ d'un Jury composé d'opérateurs humains ne sait pas dire s'il parle à un humain ou à une machine pendant cinq minutes, le test est validé. Les capacités et la puissance de calcul des premiers ordinateurs ne permettent pas de confirmer les théories et algorithmes en les implémentant mais dans les années 1980, l'évolution des ordinateurs facilite ce travail. Les premiers succès de programmes d'IA démontrent l'intérêt de ces approches. Il faut attendre les années 2010 et l'avènement du Big Data avec des volumes de données énormes ainsi que les puissances de calcul impressionnantes pour que l'IA prenne toute sa place et commence à délivrer des résultats vraiment utiles et parfois étonnants.

Ces nouvelles capacités du numérique à effectuer un travail plus rapidement, plus précisément, plus efficacement et sur des volumes de données qu'un humain a du mal à appréhender délivre de manière convaincante l'impression qu'il s'agit effectivement du travail d'une intelligence, une intelligence programmée par l'humain mais portée par la machine, bref une authentique *intelligence* artificielle. Selon nous, en l'état actuel des connaissances et des développements techniques, les systèmes d'IA ne peuvent pas (ou pas encore) être qualifiés comme d'authentiques intelligences artificielles : en dépit du vaste domaine de connaissances et de la puissance de calcul aujourd'hui disponibles (formes évoluées du traitement d'informations à grande vitesse), il manque aux systèmes d'IA des compétences d'interaction, notamment celles qui ne passent pas par le pur calcul rationnel, mais par l'émotionnel, le relationnel, l'intuition. Depuis 2014, date symbolique parce qu'elle correspondait au 60e anniversaire de la mort d'Alain Turing, la science se rapproche sans cesse de la réussite au test qu'il a proposé. Cette année-là, des chercheurs de l'Université de Reading au Royaume-Uni ont même avancé que le test avait été franchi avec succès par un programme informatique : « Eugene Goostman » a réussi à se faire passer pour un adolescent de 13 ans, originaire d'Odessa, en Ukraine. Ce programme a convaincu plus d'un tiers des juges avec qui il a discuté pendant presque cinq minutes qu'il était humain².

Mais même dans l'attente du probable succès prochain d'une intelligence artificielle au test de Turing, les questions relatives à l'intelligence des machines demeurent aujourd'hui entières. Tout se passe même comme s'il existait en la matière un véritable cap à franchir pour les dispositifs techniques, difficulté que l'on peut illustrer avec l'imagination d'un autre test. Il est permis de faire l'hypothèse que l'IA pourra être dite véritablement intelligente lorsque, compte tenu de ses capacités intrinsèques de rapidité et de profondeur dans le traitement des données, elle sera de plus en capacité d'interagir avec succès avec un enfant autiste, avec une personne âgée sénile ou handicapée psychique – test redoutable pour nombre d'intelligences humaines. C'est précisément dans sa difficulté propre que réside la valeur de cette expérience : parce qu'il suppose un niveau de complexité très élevé en termes d'interaction avec une intelligence naturelle complexe qui n'exprime pas de manière spontanément claire et distincte ce qu'elle ressent, le « test de l'humain en souffrance » pourrait même être considéré comme le test de Turing revu et corrigé, adapté à une société faisant droit à l'importance de la vulnérabilité

² Le résultat n'a toutefois pas été homologué officiellement. Voir https://www.lemonde.fr/sciences/article/2014/06/09/un-ordinateur-reussit-le-legendaire-test-de-turing_4434781_1650684.html

humaine et, dans de telles conditions, adéquat pour valider que l'IA est effectivement intelligente au sens complet du terme.

Toujours est-il que l'humain a toujours eu besoin d'appréhender son environnement à travers ses sens et son intuition, d'analyser les informations recueillies, de prendre des décisions et d'agir en retour sur son environnement avec sa parole, ses mains, son corps, en d'autres termes de se développer par des boucles d'interaction avec son milieu. Or, aujourd'hui, c'est précisément de cette manière que les applications de l'IA pénètrent, parfois en catimini, notre quotidien professionnel et personnel. Nous pouvons classer cette émergence selon ces trois dimensions des boucles d'interactions :

- Perception de l'environnement : Vision par ordinateur, reconnaissance vocale, acquisition d'information par capteurs proprioceptifs et extéroceptifs ...
- Traitement et décision : Compréhension et traitement du langage naturel, traitement et interprétation d'images, représentations et modélisation des connaissances, apprentissage supervisé et non supervisé, raisonnement, aide à la décision... avec des applications telles que recommandations/pub, filtres email spamming, assistant intelligent, conduite automatique, maintenance prédictive ...
- Capacité d'exécution : Synthèse vocale, robotique, actionneurs ...

Pour permettre la décision automatique et réfléchie de la machine, son « éducation » est clé. L'apprentissage peut être supervisé quand l'humain prépare et qualifie les données qui nourriront cette réflexion. L'apprentissage peut également être non supervisé quand les données sont fournies à la machine sans explication et que cette dernière les analyse avec des règles d'interprétation pour définir sa propre classification. Dans ces deux cas, le risque de biais introduit par l'humain est lié soit à la qualification des données en entrée soit aux règles d'interprétation proposés à la machine.

De plus, dans cette recherche de l'automatisation, le paradoxe de Moravec [Moravec 1988] nous a rattrapé : « Les tâches qui paraissent ultra simples et instinctives pour les humains sont souvent les plus difficiles à faire réaliser à une machine ». Et pourtant, combien de robots ont d'ores et déjà remplacé avantageusement l'humain sur des chaînes de production ou dans des milieux dangereux, combien de caméras permettent de voir presque mieux que Superman, combien de systèmes d'aide à la décision apportent un concours décisif à l'humain ?

La Cobotique ou la collaboration humain-robot, a quant à elle déjà réussi à associer l'humain et le robot pour tirer le bénéfice de leur complémentarité. L'Automatisation Robotisée

de Processus (*Robotic Process Automation*) permet de partager le travail efficacement entre l'humain et la machine... La répartition du travail entre l'humain et l'IA peut se faire de façon intelligente !

Dans les apprentissages humains, il y a une interaction constante entre les structures innées de pensée et de perception et l'expérience. Une IA a souvent besoin d'apprendre (*Machine Learning*) et parfois d'apprendre profondément (*Deep Learning*) pour arriver à des performances acceptables. Si l'on montre trois images de chatons à un enfant, il pourra reconnaître un chaton quand il en verra un. Et si un des chatons est dans un arbre, l'enfant pourrait même en déduire au passage que le chaton sait grimper aux arbres. Une IA a besoin de milliers d'images de chatons dans toutes les situations pour savoir identifier un chaton dans une image. La qualité de l'apprentissage supervisé ou non-supervisé conditionnera la qualité des résultats. Montrez un chat à un enfant en lui disant que c'est un chien et il le croira – du moins un certain temps. Si les données ou les règles d'apprentissage comportent des biais intentionnels ou non, l'IA se développera avec ces biais et ses résultats seront le reflet de cette « naïveté ».

2. Le coaching, une pratique sociale combinant « humanisme » et « performance » :

Il est possible d'appréhender le coaching en fonction deux filiations généalogiques différentes, l'une ancienne l'autre contemporaine.

La première est d'origine philosophique, et on peut la qualifier d'humaniste. Elle ramène en effet le propos du coaching à la tradition philosophique, les coaches aiment d'ailleurs souligner que leur activité a été initiée par le fondateur de la philosophie, Socrate³. On peut affirmer plus généralement que la référence à la philosophie antique constitue en effet pour le coaching une référence pertinente en ceci qu'elle a développé à la suite du philosophe athénien un style particulier. Parce qu'il associe étroitement la pratique de la pensée réflexive, le dialogue avec un maître de pensée et l'insertion de l'humain dans le monde via ses différentes activités [voir Hadot 2001], ce style est celui dans lequel la pratique contemporaine du coaching se reconnaît. Cette manière de pratiquer la philosophie n'a pas été seulement des philosophes de l'Antiquité, puisque, tout au long de la modernité, on la voit réapparaître chez des auteurs tels

³ <https://as-co.fr/news/socrate-precuteur-du-coaching/>

que Montaigne, Rousseau et Foucault. Les *Essais* du premier (1595) constituent une tentative pour réfléchir sa propre vie en fonction des apprentissages de sa propre expérience ; et pour les deux autres auteurs, on peut relever leur volonté de s'inscrire dans une philosophie où l'humain est invité à se regarder lui-même avec sincérité dans le contexte de son existence (thème dominant les *Confessions* de Rousseau (parus en 1782) comme les derniers cours au Collège de France de Foucault de 1983-1984 [Foucault 2009]). Cette première filiation peut être qualifiée d'humaniste, dans la mesure où les objectifs poursuivis par le coaching relèvent de l'ambition d'atteindre, pour la personne qui en bénéficie, le meilleur développement humain possible. La relation d'accompagnement ainsi caractérisée vise à faire parvenir le coaché à l'autonomie personnelle et le mettre en capacité de s'acquitter de manière heureuse de ses différentes tâches.

Apparu dans les années 1960 aux USA, le coaching peut également être compris en fonction d'une autre généalogie, celle qui interprète plus précisément son contexte intellectuel et culturel d'émergence. On perçoit alors d'autres éléments qui le caractérisent, comme son efficacité en termes d'insertion dans l'organisation et d'augmentation des performances du coaché. Du fait de son contexte d'émergence, le coaching entretient des liens avec la cybernétique et avec les éléments qui ont accompagné le développement de ce courant de pensée au-delà de l'informatique proprement dite : à savoir, les conférences de la Fondation Macy, puis l'école de Palo Alto autour des travaux de Bateson et de Watzlawick, enfin la mouvance « Nouvel Âge » [Rappin 2013]. Restituer cette origine culturelle permet de comprendre le lien qu'il entretient dans sa pratique avec les notions d'entropie, d'information et de *feedback*⁴, ainsi que son positionnement comme savoir-faire au service des organisations demandant à leurs employés de l'adaptation et de la performance.

Fort de cette double filiation, le coaching s'est répandu de par le monde pour gagner ses lettres de noblesses sous l'effet des résultats qu'il a obtenus. Le coaching se définit comme un accompagnement ponctuel d'un individu ou d'un groupe en fonction d'une demande personnelle ou dans un cadre professionnel, destiné à permettre une prise de conscience, un changement de posture, et dans le but d'amorcer et/ou de mettre en œuvre un changement et de mieux traverser une situation difficile.

⁴ Voir Rappin 2013, p. 171 : Le coaching « n'est pas simplement l'accompagnement en vue du développement du potentiel et des compétences dans le cadre d'objectifs professionnels, comme l'annoncent les ouvrages de praticiens et les associations de coaching. Ses racines cybernétiques font de lui un processus plus subtil fondé sur le lien entre information et organisation. Ainsi, le coaching est une forme d'accompagnement visant à lutter contre l'entropie organisationnelle par l'introduction d'éléments informationnels extérieurs spécifiques (le coach, sa méthodologie et ses outils). Il s'agit, sur le plan ontologique, d'un dispositif dont la finalité est la transformation de l'homme en être informationnel et la création, chez le coaché, d'une *disponibilité informationnelle* propre à « l'âge de l'accès ». ».

Plusieurs associations représentent les coachs avec notamment ICF⁵, EMCC⁶, SF Coach⁷. Née en 1995, ICF est la première fédération mondiale de coachs professionnels, avec plus de 38 000 adhérents, dont la moitié sont certifiés, répartis sur 145 chapters dans 75 pays. Tel que le définit ICF, « le coaching est une alliance entre le coach et ses clients dans un processus qui suscite chez eux réflexion et créativité afin de maximiser leur potentiel personnel et professionnel. »⁸

Si les séances de coaching peuvent être individuelles ou collectives, la nature de la relation entre le coach et le coaché est toujours la même et se voit encadrée par un contrat de coaching. Schématiquement restituées, les trois dimensions du coaching sont le « pour quoi », le « quoi » et le « comment », qui correspondent respectivement au sens, au contenu et au processus. Le coach est garant du processus, le coaché apporte le contenu, quant au sens, il concerne la finalité de la relation entre coach et coaché. Comme tout contrat, la relation de confiance qui s'établit conditionne l'effectivité des séances. Dans la construction de cette confiance, le coach apporte sa formation, ses méthodes et outils et surtout son écoute. La perception du coaché fait appel à de multiples compétences :

- Ecouter et comprendre les paroles sans jugement ni interprétation ;
- Déceler, savoir capter et gérer les émotions, les siennes et celles du coaché ;
- Faire preuve de créativité car toute personne est unique ;
- Savoir manier l'humour pour détendre ou passer des messages ;
- Utiliser son intuition quand les mots ou les faits ne sont pas suffisants ;
- Se faire confiance, avoir la conscience de soi pour diriger son attention à bon escient et intégrer tout l'environnement.

Cette relation nécessaire au bon déroulement du contrat est *intuitu personae*. Elle est fragile, fluctue et dépend des interlocuteurs, de leurs mots et actions à chaque instant. Elle ne se décrète pas par contrat, ne se modélise pas, ne se met pas en équation mais se vit.

3. Les propositions actuelles d'IA de coaching et les axes de développement possible :

⁵ <https://coachfederation.org/>

⁶ <https://www.emccfrance.org/>

⁷ <https://www.sfcoach.org/>

⁸ <https://www.coachfederation.fr/je-veux-devenir-coach>

L'IA semble vouée à envahir tous les domaines de notre quotidien : par exemple, les secteurs de la mobilité et de l'aménagement du territoire, de la santé et du bien-être, de la défense et de la sécurité, de l'administration publique, de l'organisation des entreprises (logistique, marketing, RH) sont en train de se transformer sous l'effet de l'augmentation de volume des données qui nourrit une puissance de calcul algorithmique accrue. Il faut de plus élargir ce sens restreint de l'IA, en considérant le double apport du numérique et de la robotique qui dans bien des domaines amplifie les transformations en cours et à venir. Enfin, le développement de l'Internet des objets (*IoT*) fait accéder la vie domestique quotidienne au même type de transformations. L'existence humaine et les pratiques sociales se voient davantage assistés par des systèmes apprenants, ce qui modifie la partition entre l'action proprement humaine et celle des machines.

S'il nous paraît intéressant de réfléchir aux apports possibles de l'IA dans le domaine du coaching, c'est du fait de l'émergence actuelle de dispositifs techniques dans ce domaine. Ils sont présentés comme des solutions machiniques qui permettraient aux usagers de bénéficier d'un accompagnement personnalisé plus ou moins assisté. De manière prospective, la Commission IA et Coaching de ICF a effectué une étude sur ce sujet⁹ auprès de startups et d'entreprises spécialisées dans l'accompagnement. Sa conclusion provisoire est que les entreprises concernées n'utilisent pas encore des techniques d'IA.

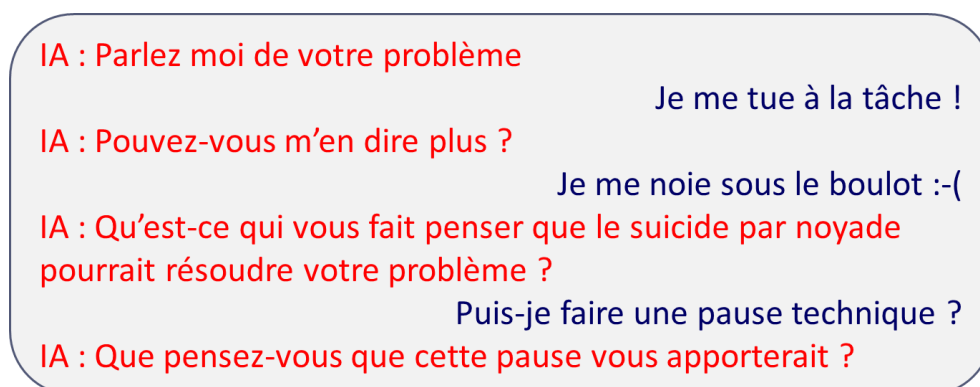
Cependant, deux ordres de raisons conduisent à pousser plus avant la réflexion. D'une part cette émergence engendre de fortes réactions tant du côté de l'informatique que du coaching, qui se traduisent par une vague d'émotions contradictoires notamment exprimées sur les réseaux sociaux : des informaticiens affirment vigoureusement que les systèmes proposés ne relèvent pas de l'IA à proprement parler, de nombreux coaches expriment un rejet spontané de ces solutions, en arguant de la perte de sens qui découlerait d'une « relation » entre une machine coach et un patient humain coaché. D'autre part, l'apport du digital et des techniques d'IA reste possible et peut même sembler parfois pertinent pour le coaching.

Le problème que nous examinons dans cette contribution se présente comme l'effet particulier d'une tendance qui affecte d'autres domaines voisins : depuis les succès inattendus du programme ELIZA¹⁰, la psychothérapie s'est depuis de nombreuses années confrontée à des

⁹ <https://www.coachfederation.fr/icf-france/commissions/intelligence-artificielle>

¹⁰ ELIZA est le nom d'un programme inventé au milieu des années 1960 par l'informaticien Joseph Weizenbaum (1923-2008), en réponse au défi lancé par Turing avec son test. Il simule les réactions d'un psychothérapeute qui serait inspiré par la méthode de Carl Rogers, en reformulant sous forme de questions la plupart des affirmations du « patient ». En dépit de ses limites techniques et du caractère sommaire de son interface, parce qu'il crée

situations comparables à celles pouvant survenir pour le coaching dans les prochaines années ; la psychiatrie se pose actuellement des questions du même type que celles que nous évoquons ici [voir par exemple Aimé 2020]. Le fait est qu'aujourd'hui de nombreuses plateformes de coaching en ligne ont vu le jour dernièrement¹¹. Elles consistent à proposer le meilleur coach selon les besoins exprimés directement par les usagers-consommateurs-clients. Les critères de recherche peuvent être plus ou moins sophistiqués mais une fois que le choix est fait, les séances de coaching s'effectuent à distance via des outils de vidéoconférence (Skype ou autres) ou le téléphone. D'autres plateformes proposent de remplacer le coach humain par des agents conversationnels (*chatbots*)¹² utilisant le traitement de la langue naturelle et certaines techniques d'IA. L'objectif, qu'on peut estimer louable dans son principe, consiste à permettre l'accès au coaching au plus grand nombre. Ces « dialogueurs » suivent le processus de coaching et posent des questions prédéfinies en fonction des informations ou réponses fournies par le coaché. L'analyse syntaxique sert à identifier les mots clés sur lesquels le *chatbot* rebondit. Si la conversation peut aider le coaché à prendre du recul et à se poser les bonnes questions, le fait que les questions soient prédéfinies ne permet pas un travail sur l'analyse sémantique. Ceci est dû à la compréhension encore limitée de ces *chatbots* notamment sur le sens et la valeur des mots dans le contexte de la discussion. On peut fournir quelques exemples possibles de malentendus possibles en restituant ce dialogue fictif entre un humain coaché et un IA « coach » :



IA : Parlez moi de votre problème
Je me tue à la tâche !

IA : Pouvez-vous m'en dire plus ?
Je me noie sous le boulot :-)

IA : Qu'est-ce qui vous fait penser que le suicide par noyade pourrait résoudre votre problème ?
Puis-je faire une pause technique ?

IA : Que pensez-vous que cette pause vous apporterait ?

Figure 1 : exemples de malentendus possibles dans le dialogue entre une IA coach et une personne humaine coachée

l'illusion d'un dialogue avec l'humain sur des questions de nature privée, ELIZA s'est avérée susceptible de créer une forte dépendance émotionnelle.

¹¹ <http://www.international-coaching-solutions.fr/coaching-en-ligne/> ; <https://www.coachready.com/fr/accueil/> ; <https://moovone.eu/plateforme-coaching-en-ligne/> ; <https://outilsducoach.com/coaching-ligne/>

¹² <https://pocketconfidant.com/> ; <https://replika.ai/>

Par ailleurs, si cet apport encore très limité apparaît pertinent dans son principe, c'est dû aussi bien aux limites humaines qu'au potentiel des technologies. Ainsi, malgré tout le professionnalisme du coach lors d'une séance, son attention peut se trouver distraite et le coach peut manquer d'attention à certains messages verbaux comme non-verbaux. Une IA qui se constituerait comme une véritable intelligence capable d'interactions avec l'humain ne connaîtra pas de tels aléas et sera en mesure de garder sa concentration maximale sans aucune émotion ou sentiment parasite. Nous pouvons tenter de dresser la liste des thématiques et des techniques qui peuvent déjà ou pourront demain apporter des plus appréciables à la pratique du coach :

- Sous réserve de l'accord du coaché, l'enregistrement des séances de coaching peut permettre une analyse sémantique pour extraire un corpus, faire ressortir les mots qui reviennent où sur lesquels le coach devra travailler. L'intonation de la voix, les hésitations peuvent également être analysées pour proposer des relances ou des focus sur des points particuliers
- Toujours sous réserve de l'accord du coaché, une caméra et des capteurs physiologiques peuvent renseigner sur les états internes et tensions émotionnelles du coaché
- A partir des thèmes abordés, le système peut aller chercher des informations complémentaires, des sources et références, des statistiques, des cas similaires...
- Le suivi des actions posées entre les séances de coaching peut être accompagné de façon automatique par des relances et des encouragements
- La prise de rendez-vous et leur report sont possibles de façon automatique
- Enfin, un outil de supervision de coach peut permettre un retour et un debriefing à chaud avant d'en parler à son superviseur...

Toutes ces aides informatiques viennent en appui et complément des actions du coach sans le remplacer. Si nous devions comparer les avantages et inconvénients d'une IA et d'un coach humain, la balance pencherait certainement du côté de l'IA pour la disponibilité 24/7, l'égalité de l'humeur, la collecte d'information et la recherche de références.

Les points qui sont sujets à discussion regardent les aspects suivants :

- La neutralité : si l'humain est influençable et ne peut s'empêcher d'interpréter et de juger, même un coach, l'IA devrait être neutre ... sauf si son apprentissage a été biaisé par des jeux de données biaisées.

- Le respect de la vie privé : est-ce qu'un système dans le cloud mais certifié RGPD est plus fiable qu'un humain ?
- La relation coaché – coach : créer la relation, libérer la parole, bâtir la confiance nécessaire à s'ouvrir à l'autre... est-on plus à l'aise avec une personne ou en parlant à une machine ?
- La connaissance du contexte : autant une IA pourra accéder à un savoir quasi infini, autant seul un coach saura quel savoir est pertinent. Sur cet aspect, la disparition du tiers humain dans la relation de coaching s'avère évidemment préjudiciable à la bonne effectuation de l'activité.

Reflète du paradoxe de Moravec précédemment évoqué, il semble que l'IA connaîtra de grandes difficultés dans les situations dans lesquelles le bon sens commun est nécessaire. Si l'IA peut englober des montagnes de connaissances explicites, les connaissances tacites et leur compréhension selon le contexte passent difficilement et risquent de laisser un arrière-goût amer. Une autre dimension est susceptible de mettre l'IA en grande difficulté, à savoir, celle où ce que l'on dit n'a pas le sens premier. Les traits d'humour, le deuxième degré, l'autodérision, les sous-entendus, les métaphores, les analogies, et jusqu'aux non-dits que l'on peut deviner si on est humain ! Ainsi que le rappelle le psychiatre Thierry Vincent, si les machines excellent à remplir des grilles de *sudoku*, seuls les humains savent faire des mots-croisés, dans lesquels les jeux sémantiques font accéder à la signification des mots en fonction de leurs multiples sens [Vincent 2019 : 178].

Thématiques	Coach	IA
Respect de la vie privée		
Disponibilité 24/7		
Etat interne du coach		
Relation entre coach et coaché		
Humour, 2 ^e degré, auto-dérision, non-dits		
Utilisation de métaphores et d'analogies		
Décisions/déductions rationnelles		
Statistiques et sources d'informations		
Coaching paradoxal		
Piège abscons, conflit interne, manipulation, biais		
Comportements irrationnels		
Connaissance du contexte		
Réconfort, contact physique		



Figure 2 : un exemple des points de discussion actuels entre coachs et IA

Le tableau ci-dessus illustre la difficulté que chacun aurait à donner sa confiance à un interlocuteur humain ou machine selon les thématiques.

Enfin, et même si la déontologie a son mot à dire sur ce point, le réconfort physique que le coach peut apporter par son attitude et le non-verbal sera inaccessible à une machine ou un robot...du moins, pour l'instant, tant les aspects sensibles de la robotique et même de l'IA sont susceptibles de faire concurrence aux humains dans la dimension de la relation. Ainsi, certaines œuvres de fiction, très populaires, ont déjà souligné les relations d'attachement, de transfert et de projection qui peut nourrir le lien entre l'humain et la machine par l'intermédiaire de simulations sensorielles telles que le son d'une voix qui semble humaine (voir à cet égard le très crédible film *Her* de Spike Jonze, 2013). A la fois compte tenu du nombre d'artefacts déjà à l'œuvre au milieu de l'activité humaine et prenant acte de l'émergence d'interfaces sensorielles de qualité, les spécialistes du comportement humain indiquent qu'un « nouveau pacte » entre l'humanité et les machines est probablement déjà en train de se mettre en place [Vidal 2012], tandis que les chercheurs en robotique peuvent déjà compter sur la constitution d'une authentique « sociologie » dans la relation avec les machines [Borelle 2018]. En tout état de cause, la *personnalisation* par les usagers des machines qui sont déjà à leur disposition atteste également d'une tendance à leur *personnification* : par exemple aujourd'hui avec les smartphones, les ordinateurs portables, mais aussi (sous l'effet d'un très habile marketing) les enceintes connectées à reconnaissance vocales – et demain sans doute l'ensemble de « l'internet des objets » ou *IoT*.

Cette tendance semble en partie naturelle en partie créée par le design contemporain, en tout cas elle apparaît inquiétante à plusieurs titres. Premièrement, elle inquiète parce qu'elle abolit la barrière entre l'humain et ses artefacts ; deuxièmement, parce qu'elle s'opère *via* le registre émotionnel ou infra-rationnel, et également en entremêlant le rationnel et l'émotionnel ; troisièmement, parce que la projection émotionnelle envers les machines s'avère asymétrique : les humains investissent de sens les dispositifs techniques à travers des émotions que, par essence, ces dispositifs n'éprouvent pas pour leur part – la relation entre l'humain et la machine, qui constitue pourtant une « boucle affective » dans lequel le premier nommé trouve le sens de ses pratiques avec les machines, apparaît de ce fait entièrement asymétrique [voir à ce sujet Dumouchel et Damiano 2016 : 149-157] .

4. Déontologie du coaching et éthique de l'IA :

Dans ces conditions de bouleversement et de confusion, l'appel à la dimension éthique s'avère nécessaire. Nous entendons par « éthique » aussi bien la recherche d'un sens pertinent pour l'action que la détermination de règles conformes à des valeurs considérées comme justes et bonnes. Or, sur cet aspect, nous voulons souligner que, loin d'être figée, la situation peut évoluer dans plusieurs directions différentes : si l'IA peut apporter quelque chose au coaching, la réciproque est envisageable, et cela sur deux aspects.

En premier lieu, le coaching aura à traiter tous les nouveaux problèmes générés par l'IA... car les progrès technologiques contribuent à faire monter l'inquiétude de la compétence professionnelle des humains en regard des machines : « Je vais perdre mon job », « un robot va me remplacer », « une IA est plus intelligente que moi » ... Certes, ces questions ne sont pas nouvelles, mais aujourd'hui l'adversaire apparaît réellement redoutable. Comment se battre contre un concurrent qui peut soulever 100kg d'une main, qui peut répéter un même geste 1000 fois avec une précision « diabolique » (sur les robots et le diable, voir [Grinbaum 2018]), qui peut voir à travers les choses, qui peut calculer à la vitesse de la lumière ou se souvenir de tout... ? Accompagner les personnes qui seront confrontées à ces questionnement demande une approche peut-être différente...ou tout du moins, de revoir l'argumentaire.

Dans un deuxième lieu, la notion d'IA augmentée par le coaching avec ce que l'« état coach » pourrait apprendre à l'IA. Il faut rappeler combien le « *Primum non nocere* » cher aux médecins tient également à cœur aux coachs, logiquement soucieux de la démarche éthique [pour une mise au point récente sur les relations entre la pratique du coaching et l'éthique de la profession, voir Poulet 2019]. Le coaching représente ainsi une activité complexe qui mobilise les ressources du coach de manière globale, si bien que, comme l'écrit Poulet, « le ou la Coach doit se mobiliser de manière holistique lorsqu'il ou elle accompagne ses clients : au niveau corporel, émotionnel et cognitif ». La conscience de cet engagement conduit le praticien à une double recherche permanente, celle de règles pour encadrer la pratique et celle d'une ligne de conduite claire pour lui permettre de parvenir à un résultat satisfaisant. Dans les termes choisis par Poulet, la *déontologie*, produite et garantie par le cadre d'exercice de la profession qui correspond aux grandes fédérations de coaching, représenterait son cadre extérieur, et *l'éthique* sa dimension intérieure (« [...] la déontologie apparaît comme un cadre moral et comportemental extérieur, imposé au coach. L'éthique est, elle, intérieure et propre à chaque personne. »). Cette partition de base permet d'engager, au sein de la relation entre le coach et

le coaché conçue comme une « zone de travail », la recherche permanente d'une forme de justesse, en conformité avec « la morale » qui représente selon Poulet la conformité que les personnes entretiennent avec leurs propres valeurs.

A la lumière de ces éléments, on voit bien ce qui peut se trouver transformé par l'apparition de formes machiniques de suivi et d'accompagnement. Jusqu'ici, les valeurs de bienveillance et de respect de l'humain sont l'apanage des humains. Tandis que ceux-ci, ainsi qu'on l'a relevé plus haut, ont tendance à se relier à leurs dispositifs techniques par des projections affectives, comment réussir à incarner dans la machine la bienveillance et le respect de l'humain, et à les coder dans les processus d'apprentissage des IA ? La question serait alors : comment transcrire une éthique en programme informatique, avec quelles règles et quels apprentissages ?

Mais si l'on prend les choses en rigueur de termes, ce sont des séries de questions assez différentes qui apparaissent. Ou bien on estime que le dispositif technique est éthiquement neutre par lui-même, et qu'il ne fait que refléter les valeurs des concepteurs (eux-mêmes engagés dans la voie d'une tentative pour que les algorithmes soient le plus transparent possible). En ce cas, le questionnement éthique se trouve déporté sur la responsabilité des usagers¹³. Ou bien – et en dépit du fait que l'on ne parle pas ici de dispositifs techniques réellement dotés de réflexivité donc moralement autonomes comme le sont les humains –, on estime que la machine doit être en mesure de développer une éthique ou des contenus éthiques. En ce cas, il serait nécessaire de coder les programmes de manière conforme aux grandes éthiques, et non de se limiter aux formes actuelles de la *computer ethics*, dont la forme massivement admise aujourd'hui par les informaticiens se borne à une approche particulière, qui est typique de l'utilitarisme ou « conséquentialisme » et qui vise à caractériser des dilemmes permettant un choix de valeur satisfaisant.

Toujours est-il qu'il apparaît pertinent de tenter de pluraliser les formes du raisonnement éthique, au lieu de s'enfermer dans le seul paradigme conséquentialiste. La littérature de philosophie morale offre justement la ressource de trois autres formes possibles de raisonnement éthique : au conséquentialisme peut en effet se trouver associé le déontologisme (ensemble des éthiques qui privilégient l'expérience morale intériorisée ou devoir), l'arétaïsme (terme sous lequel se regroupent les éthiques des vertus) et l'axiologisme (ensemble des

¹³ Tel semble être le point de départ de l'expérimentation du MIT relative au véhicule autonome, voir <http://moralmachine.mit.edu/hl/fr> et [Awad 2018]

éthiques privilégiant une valeur cardinale, on peut ici convoquer les éthiques d'origine religieuse). Le déontologisme peut être illustré par la figure de Kant qui, dans les *Fondements de la métaphysique des mœurs* (1785) et dans la *Critique de la raison pratique* (1788) s'est attaché à formuler la législation éthique qui conditionne la capacité humaine à forger un jugement moral autonome, c'est-à-dire un jugement basé sur la faculté d'orienter sa propre action par une réflexion désintéressée. Pour illustrer l'arétaïsme, on peut évoquer Aristote ou Augustin d'Hippone, des auteurs pour lesquels peut être dite éthiquement valable la décision ou l'action qui à la fois *repose sur* et *permet de développer* les vertus, à savoir, des éléments de caractère socialement féconds tels par exemple que le courage, la justice et la tempérance pour Platon, la magnanimité, la libéralité, selon Aristote, ou encore la bonté pour le christianisme (tout en disqualifiant les vices, à savoir, les éléments de caractère socialement néfastes tels que la lâcheté, l'avidité, l'impulsivité, l'intolérance, l'avarice, etc.). Enfin, l'axiologisme peut renvoyer à des discours fondamentalement différents les uns des autres : par exemple, les morales religieuses, pour les monothéismes, la volonté d'un Dieu unique constitue en ce cas la valeur suprême, et on peut l'illustrer avec la figure de leurs prophètes ; par exemple encore, les éthiques philosophiques prônant des valeurs supérieures comme celle de la nature (des auteurs comme Hans Jonas ou Arne Naess apparaissent comme de bons candidats pour l'éthique écologiste) ou celle de la nation ou du vivre-ensemble (Ernest Renan ou Hannah Arendt).

En dépit du tropisme de l'éthique computationnelle dominant actuellement le paysage de l'IA, ce logicisme qui privilégie le conséquentialisme d'origine utilitariste, ce n'est pas une seule, mais bien quatre formes d'éthique qui existent. Toutes sont cohérentes, attestées dans la littérature, et apparaissent pertinentes pour la démarche visant à enrichir la contribution humaine aux choix éthiques face aux développements de l'IA. Pour examiner les relations entre l'IA et le coaching dans une visée d'évaluation éthique, il serait donc sans aucun doute aussi utile que pertinent de formuler des arguments empruntés non pas à une seule, mais bien aux quatre familles de l'éthique. D'ailleurs, le codage d'un dispositif artificiel de coaching en fonction des thématiques typiques de chaque famille de l'éthique n'est pas impossible et pourrait représenter une avancée majeure pour garantir le développement d'une IA de coaching considérée comme éthique.

Conclusion :

On a vu comment, dans sa contribution citée plus haut, Baptiste Rappin a établi une généalogie du coaching à partir de la cybernétique – quelles peuvent être les conséquences de

la restitution sur le sujet qui nous préoccupe ? Un passage particulier attire notre attention, qui souligne l'importance dans ce modèle de la circulation de l'information :

« [...] Si, du point de vue ontologique, l'organisation résulte de la circulation de l'information, et si, du point de vue politique, le feedback permet de s'assurer du contrôle de cette circulation, alors il faut créer les conditions éthiques, c'est-à-dire effectives chez l'être humain, d'une telle circulation : à savoir faire de la circulation de l'information une valeur en soi susceptible d'être incarnée. Le coach est précisément le modèle de cette incarnation, lui chez qui la technique et l'axiologique se rejoignent pour se confondre (par exemple, le dialogue comme technique ET comme valeur). »
[Rappin 2013 : 174]

Dans cette perspective, on pourrait dire que l'éthique du coaching, au sens de : l'éthique la mieux adaptée à cette pratique particulière, consisterait à augmenter la quantité d'information disponible et à faire circuler l'information du mieux possible, afin d'accroître l'intelligence que le coaché a de sa propre situation et de son insertion dans son cadre professionnel. De ce point de vue, le rapprochement entre la pratique du coaching et sa filiation cybernétique donne à imaginer ce que serait l'apport de l'IA sur le coaching. Car pour certains développements récents de l'éthique de l'informatique également, une définition pertinente de l'éthique consiste en ces deux termes : l'accroissement de l'information disponible et la circulation de l'information la meilleure possible, dans l'idée que cette perspective permet d'offrir au système apprenant le cadre adéquat à son perfectionnement. Ainsi que l'explique Grinbaum, pour la machine, « une information nouvelle est un bien » [Grinbaum 2018 : 189]. Si l'IA est ainsi conçue, son apport à l'éthique du coaching consisterait à concevoir des systèmes d'accompagnement réellement ou fortement interactifs, qui mettent en jeu des données à la fois riches et transparentes. Et dans ce cadre, d'après l'état actuel des connaissances sur l'aspect « incarné » de la relation humains-machines [Damiano & Dumouchel 2019], une palette émotionnelle enrichie et la plus compréhensible possible contribuerait elle aussi à une amélioration des dispositifs proposés.

Au vu de ces éléments, en dépit du caractère apparemment iconoclaste d'une telle position, nous estimons enfin que les coachs, du fait de leur expertise en matière de relations humaines, pourraient gagner à participer activement à cette évolution en collaborant avec les ingénieurs développeurs. D'une part, en effet, leur participation s'avère nécessaire pour préserver les valeurs et l'éthique du coaching – aujourd'hui basées en grande partie sur la présence de l'humain dans la relation, à la fois en tant que personne capable de sensibilité et de bon sens et en tant que tiers. De l'autre, l'aventure contemporaine de l'invention de systèmes

machiniques de coaching leur offre l'opportunité d'appréhender leur pratique de manière originale et renouvelante. La ligne de conduite à tenir pour que cette aventure ne tourne pas mal peut être fournie par une actrice-témoin de l'innovation technologique, en la personne de Judith S. Donath, fondatrice du *Sociable Media Group* au « MIT Media Lab » du Massachusetts Institute of Technology et pionnière des applications de média social. Dans un article mettant l'accent sur l'importance d'« être réel » (*Beeing real*) pour réussir une communication humaine, et en revenant sur le cas d'ELIZA, Judith S. Donath met en garde contre les « télé-identités » qui risquent, via les applications numériques, d'engendrer une « télé-éthique » vide de sens :

« Enfin, il est important de garder en tête que le but de la communication en grande partie n'est pas l'échange d'informations factuelles, mais l'établissement et le maintien de liens sociaux et de structures. Nous communiquons au travers d'un support, pour acquérir un statut, se faire des amis. Dans ce cas, l'identité de nos compagnons – et certainement leurs qualités humaines – est de première importance. Weizenbaum, le créateur d'ELIZA, fut horrifié quand son programme fut accueilli, non pas comme un reproche fait à l'équation de Turing, disant qu'on agit avec intelligence en étant intelligent, mais comme un compagnon divertissant ou un précurseur de la psychothérapie automatisée. Pour Weizenbaum, cet enthousiasme était « déplacé de manière obscène », une menace directe pour notre humanité. [...] Dans la mesure où le monde virtuel croît jusqu'à englober tous les aspects de nos vies et de nos interactions en ligne, façonner nos communautés, influencer nos opinions politiques et servir d'intermédiaire avec nos proches, dès lors la qualité d'être réel, qui est acceptée et présumée de fait dans le monde physique, devient une des questions centrales de société. » (Donath 2003 : 47]

Le défi qui se pose au coaching assisté par l'IA peut donc être formulé ainsi : savoir demeurer « réel » grâce à la qualité de la relation avec le coaché, appréhendée comme un lien qui fait sens dans un environnement social en transformation et qui appelle la prise de points de repère.

Références :

- Aimé (Xavier), “Intelligence artificielle et psychiatrie : noces d'or entre Eliza et Parry”, *L'information psychiatrique*, 2017/1, Volume 93, p. 51-56, consulté le 24/03/2020 à l'URL : <https://www.cairn.info/revue-l-information-psychiatrique-2017-1-page-51.htm>
- Awad, Edmond, *et alii*, 2018: “The Moral Machin Experiment”, *Nature*, Volume 563, p. 59–64.
- Borelle (Céline) 2018 : « Sortir du débat ontologique. Éléments pour une sociologie pragmatique des interactions entre humains et êtres artificiels intelligents », *Réseaux*, 2018/6 n° 212, p. 207-232.

- Dascal (Marcelo) 2005 : « Culture numérique. Enjeux pragmatiques et philosophiques », *Diogène*, 2005/3, n° 211, p. 26-47, consulté le 25/03/2020 à l'URL <https://www.cairn.info/revue-diogene-2005-3-page-26.htm>.
- Donath (Judith S.) 2003 : « Être réel », *Sociétés*, 2003/1, n°79, p. 35-49, consulté le 25/03/2020 à l'URL <https://www.cairn.info/revue-societes-2003-1-page-35.htm>
- Dumouchel (Paul) et Damiano (Luisa) 2016 : *Vivre avec les robots. Essai sur l'empathie artificielle*, Paris, Editions du Seuil.
- Foucault (Michel) 2009 : *Le courage de la vérité. Le gouvernement de soi et des autres II*, Cours au Collège de France. 1983-1984, éd. F. Gros, Paris, Gallimard-Le Seuil.
- Grinbaum (Alexei) 2018 : *Les Robots et le mal*, Paris, Desclée de Brouwer.
- Hadot (Pierre) 2001 : *La Philosophie comme manière de vivre*. Paris, Albin Michel.
- Le Blanc (Benoît) 2014 : « Alan Turing : les machines à calculer et l'intelligence », *Hermès*, 2014/1, n° 68, p. 123-126, <https://www.cairn.info/revue-hermes-la-revue-2014-1-page-123.htm>
- Moravec (Hans) 1988 : *Mind Children. The Future of Robot and Human Intelligence*, Cambridge (Mass.), Harvard University Press.
- Poulet (Florent) 2019 : « La notion d'éthique dans le coaching professionnel », *Revue européenne de coaching*, numéro 8, 09/2019, consulté le 19/02/2020 à l'URL <https://revue-europeenne-coaching.com/numeros/n8-septembre-2019/la-notion-dethique-dans-le-coaching-professionnel>
- Rappin (Baptiste) 2013 : « Essai philosophique sur les origines cybernétiques du coaching », *Communication et organisation* [En ligne], 43 | 2013, consulté le 03/02/20 à l'URL <http://journals.openedition.org/communicationorganisation/4204>
- Turing (Alan) 1950 : « Computing Machinery and Intelligence », *Mind*, Vol. LIX, n°236, Octobre 1950, p. 433-460, consulté le 25/03/2020 à l'URL <https://academic.oup.com/mind/article/LIX/236/433/986238>
- Vidal (Denis) 2012 : « Vers un nouveau pacte anthropomorphique ! », *Gradhiva* [En ligne], 15 | 2012, mis en ligne le 16 mai 2015, consulté le 25/01/2020 à l'URL <http://gradhiva.revues.org/2319>
- Vincent (Thierry) 2019 : « L'intelligence artificielle du post-humain. Une crise de la raison », in Ollier (Vincent), *La transmutation posthumaniste. Critique du mercantilisme anthropotechnique*, Quel Sport ? Editions, p. 169-178.