

Une conscience phénoménale est-elle possible chez des agents artificiels ?

GENEVIÈVE FRÉCHETTE, *Université Laval*

RÉSUMÉ : Plusieurs chercheurs dans le domaine de l'intelligence artificielle veulent montrer qu'il est possible de concevoir des machines capables de reproduire le comportement humain. Certains d'entre eux vont encore plus loin et postulent que ces agents artificiels pourraient être munis d'une conscience comme la nôtre - c'est-à-dire d'une conscience phénoménale - et éprouver des sensations, avoir des expériences perceptives, ressentir des émotions, etc. Toutefois, cette thèse est-elle vraisemblable ? Dans cet article, j'analyserai la possibilité d'une conscience phénoménale chez les agents artificiels. Je mettrai en lumière ce qui me paraît être des caractéristiques essentielles d'une telle conscience, pour examiner ensuite s'il est envisageable que celles-ci se retrouvent chez les agents artificiels. Je répondrai finalement que le projet de l'intelligence artificielle forte ne me semble pas réalisable, puisqu'à première vue, un robot ne peut remplir les critères nécessaires d'une conscience phénoménale, qui exige une vie subjective que les agents artificiels ne peuvent posséder.

Introduction

Face aux progrès des sciences cognitives, plusieurs questions d'envergure émergent, notamment en ce qui a trait à l'intelligence artificielle et, plus précisément, à l'intelligence artificielle forte. Selon les défenseurs de l'intelligence artificielle forte, les processus mentaux doivent être compris comme des processus computationnels. Ainsi, il serait possible de concevoir une machine réellement consciente qui serait dotée d'une vie mentale telle que la nôtre. Il va sans dire que si l'idée d'un agent artificiel conscient était réalisable, cela bouleverserait considérablement le paradigme en place. La validation d'une telle hypothèse aurait, entre autres, pour conséquence d'amener

la nécessité de repenser la sphère de la politique, de l'éthique, du droit, de l'éducation, etc. Toutefois, plutôt que de m'écarter vers des conclusions hâtives, je désire porter un regard en amont, en posant une question qui vise l'enjeu même du problème : une conscience phénoménale est-elle possible chez des agents artificiels ? Dans cet article, je tenterai de montrer en quoi cette idée me paraît inadmissible. Pour ce faire, je mettrai en lumière ce qui semble être les critères d'une conscience phénoménale en m'inspirant principalement de la phénoménologie de Husserl. Par la suite, je me pencherai sur les prétentions de l'intelligence artificielle forte et tenterai de montrer que celles-ci sont difficiles à soutenir en raison des caractéristiques propres aux agents artificiels. En effet, ceux-ci ne semblent pas satisfaire aux conditions de possibilité d'une conscience *phénoménale*¹, ce qui me pousse à douter de la validité des thèses de l'intelligence artificielle forte.

1. *Qu'est-ce qu'une conscience phénoménale ?*

Tout d'abord, il est nécessaire de bien comprendre les termes concernés par l'argumentation qui suivra, en commençant par le concept de conscience phénoménale. Étant donné la nature mystérieuse de la conscience, un éventail de significations lui est attribué, mais je me contenterai de mettre en lumière deux niveaux de définition, soit la conscience d'un point de vue biologique et la conscience d'un point de vue phénoménal. D'un côté, la conscience biologique réfère davantage aux processus mentaux comme la perception, la mémoire, l'imagination, etc. On peut également l'appeler cognition, comprise comme l'ensemble des processus de traitement de l'information. D'un autre côté, la conscience phénoménale - nommée aussi *expérience consciente* - se qualifie plutôt comme l'ensemble des vécus d'un sujet, son expérience subjective, ses qualia². Autrement dit, quelle que soit l'expérience que je fais, la conscience phénoménale est *l'effet* que cela me fait de faire une telle expérience.

Une conscience phénoménale est-elle possible chez des agents artificiels ?

2. *Les conditions de possibilité de la conscience phénoménale*

Afin de déterminer si une conscience phénoménale est possible chez les agents artificiels, il importe d'identifier ce qui paraît être les conditions de possibilité d'une telle conscience. Dans la mesure où le discours traitant de la conscience se veut être cohérent, quels seraient ses caractéristiques essentielles et ses critères d'effectivité ? Tout d'abord, il semble que toute conscience phénoménale doit avoir un *corps matériel* comme support, en plus d'être dotée d'une libre *volonté*. Ensuite, pour être phénoménale, la conscience doit avoir le caractère de *subjectivité* et d'*intentionnalité*. L'expérience consciente doit également pouvoir faire une *synthèse de ses vécu*s et avoir le caractère d'*auto-affection*, c'est-à-dire qu'elle a comme caractéristique d'être à la fois sujet et objet. Enfin, il semble qu'une conscience phénoménale ne peut être pensée qu'avec une *temporalité* et une *spatialité* - au sens où elle est toujours un « ici » et un « maintenant ». Ces caractéristiques seront développées davantage dans les paragraphes qui suivent.

2.1. *Le corps matériel*

Une conscience phénoménale n'est pas, autant que l'on puisse en juger, imaginable sans un corps. Les deux forment un tout dont le corps est compris comme l'extension de la conscience. Si je suis en mesure de dire que l'expérience que je fais du monde qui m'entoure est toujours *miennes*, c'est grâce à *mon* corps. Je suis capable de toucher mon corps en tant que substrat matériel en même temps que mon corps *se sent* touché ; « d'une part, il est chose physique, *matière*, il a son extension dans laquelle entrent ses propriétés réelles³, la coloration, le lisse, le dur, la chaleur et toutes les autres propriétés matérielles du même genre ; d'autre part, je trouve en lui et je *ressens* "sur" lui et "en" lui : la chaleur du dos de la main, le froid aux pieds⁴ ». Le corps est le support de ma conscience du fait qu'il a les organes sensoriels qui me permettent d'être en relation avec le monde, mais aussi parce qu'il est le lieu d'apparition des phénomènes. En ce sens, je suis également convaincue que la conscience phénoménale doit avoir comme support un corps *biologique*⁵. John Searle dira que « la conscience est causée par des

processus neuronaux de niveau inférieur dans le cerveau, et est elle-même une caractéristique du cerveau. [...] Nous pouvons penser qu'il s'agit d'une "propriété émergente" du cerveau⁶ ». Ainsi, la conscience pourrait être explicable causalement par les propriétés neuronales, mais elle n'est pas pour autant réductible à la somme de celles-ci. Même si nous pouvons penser qu'il y a une relation causale entre les processus neuronaux et l'apparition de la conscience, prétendre que la conscience phénoménale puisse être comprise comme la somme des processus neuronaux est une conclusion illégitime. Autrement dit, le fait que la conscience émergerait du cerveau ne justifie pas à elle seule une thèse, comme celle du computationnalisme fort, qui réduirait la conscience phénoménale à un simple système de traitement de l'information. Or, nous pouvons très bien imaginer que l'expérience consciente est rendue possible grâce au cerveau sans pour autant prétendre que la conscience se résume aux processus neuronaux.

2.2. *La subjectivité*

L'un des traits les plus importants de la conscience est son caractère subjectif. En effet, il semble qu'une part - voire une très grande partie - de notre expérience échappe à l'objectivité. Par exemple, mon ami reçoit une bonne nouvelle qui le met dans un état de joie intense. Même s'il m'explique ce qui le rend joyeux et qu'il me décrit le plus précisément possible ses émotions, il demeure, pour moi, impossible de saisir l'essence de son vécu. L'expérience vécue est la *sienne*. Or, je peux avoir de l'empathie, imaginer comment il se sent, mais ce sera encore et toujours incomplet. De plus, même si je parvenais à comprendre parfaitement ce qu'il vit, je n'aurai pas accès à l'effet que cela *lui* fait, mais plutôt à ce que cela *me* fait, en tant que *mon* propre vécu. En somme, l'expérience de la conscience phénoménale a comme trait d'essence d'être subjective, elle est toujours pour une conscience unique, de telle sorte que nos expériences sont vécues selon notre point de vue individuel. Nagel dira qu'« il est difficile de comprendre ce que pourrait signifier le caractère *objectif* d'une expérience, indépendamment du point de vue particulier à partir duquel son sujet l'appréhende. Après tout,

Une conscience phénoménale est-elle possible chez des agents artificiels ?

que resterait-il de l'effet que cela fait d'être une chauve-souris si l'on ôtait le point de vue de la chauve-souris⁷ ? » Ainsi, l'entreprise scientifique qui veut expliquer le fonctionnement du cerveau par les voies de la physique ne parviendrait pas à expliquer la réalité d'un autre ordre, soit celle de la phénoménalité.

2.2.1. Les états intentionnels

De plus, il importe de mentionner que l'expérience consciente n'a de sens qu'au sein d'une compréhension de la conscience comme essentiellement *intentionnelle*. Ainsi, les choses apparaissent à une conscience à travers un vécu et celui-ci est intentionnel dans la mesure où il est toujours un vécu *de quelque chose*. Nous n'aurions accès à rien de ce qui nous entoure si les objets ne nous apparaissaient pas en tant que phénomènes. Or, mon accès à l'arbre devant moi n'est possible que dans la mesure où ma conscience le vise (j'ai conscience *de* l'arbre) et qu'il se phénoménalise à moi (l'arbre apparaît à ma conscience et j'ai un vécu de l'arbre). Je ne parviens jamais véritablement à la chose dans sa réalité empirique, elle n'apparaît à moi qu'en tant qu'objet de signification. Par le moyen du sens, je perçois l'arbre en tant qu'arbre et celui-ci apparaît en tant que *mon* phénomène⁸.

2.2.2. La synthèse des vécus de conscience

Cependant, le sens intégral de mon objet de perception ne m'est jamais donné du simple fait que je l'ai visé. Par exemple, quand je regarde l'arbre, je ne peux pas percevoir en même temps toutes ses faces (de côté, de l'intérieur, du dessus, etc.). Ainsi, la conscience doit procéder à une perpétuelle synthèse en vue de constituer l'unité de sens de l'objet et elle opère constamment une liaison entre les vécus de conscience en vue de parvenir à une unité⁹. Par exemple, prenons une personne qui n'a jamais vu ou entendu parler de la neige. Lorsqu'elle en fera l'expérience, elle procèdera à une synthèse de ses vécus, ce qui lui permettra de l'identifier comme froide, blanche, etc. Certes, sa synthèse du vécu de la neige n'aura pas la même complétude que la mienne - habitant au Québec depuis 28 ans - mais, dans les deux cas, la liaison de nos propres vécus

antérieurs aura pour effet d'affecter notre expérience phénoménale de la neige.

2.3. *L'auto-affection*

L'accès à une conscience phénoménale requiert un autre critère, celui de l'*auto-affection*. Avoir une conscience phénoménale nécessite d'être en rapport avec les choses du monde. Or, ce n'est pas quelque chose dont nous doutons. Par exemple, nous savons que la neige est froide puisque nous en avons fait l'expérience et si nous pouvons faire une telle expérience, c'est grâce à notre caractère d'auto-affection. Autrement dit, je suis toujours sujet et objet de ma propre expérience. Je ne peux pas me considérer simplement comme un corps ou comme n'importe quel objet biologique. En vérité, l'expérience vécue de mon propre corps révèle, de surcroît, mon expérience à la première personne, moi, en tant que sujet. Ainsi, en parvenant à l'unité de mon propre corps et de son expérience subjective, je confirme ma participation au monde. Le monde est alors toujours *mon* monde ; « le monde environnant n'est pas monde "en soi", mais monde "pour moi", c'est-à-dire justement monde environnant de *son propre* sujet égologique, monde dont le sujet fait l'expérience [...] qu'il pose au sein de ses vécus intentionnels avec la teneur de sens qui y est chaque fois impliquée¹⁰ ». On peut dire que la conscience phénoménale affecte le monde en tant qu'elle vit au sein de *ses* objectivations ; les objets du monde sont ses objets. Mais elle est également toujours affectée et transformée par eux dans la mesure où ses comportements sont déterminés par le monde¹¹.

2.4. *La spatialité et la temporalité*

Cette capacité à avoir une vision unitaire de soi en tant qu'objet et sujet me permet, en outre, de me situer dans le monde en tant que « point zéro ». Les choses du monde entretiennent un rapport à ma conscience en tant qu'elle est le lieu d'apparition des phénomènes. Je ne pourrai jamais m'éloigner de moi, je suis le « ici et maintenant ». En cela, je peux dire que les choses m'apparaissent dans un temps et un espace qui me sont propres, qui sont toujours relatifs à moi. De plus, ma temporalité se dégage du fait que mon expérience

Une conscience phénoménale est-elle possible chez des agents artificiels ?

consciente est déterminée par mes vécus antérieurs. En reprenant l'exemple de la neige, mon expérience phénoménale de celle-ci, le vécu ressenti que j'ai, dépend de mon expérience passée qui m'est propre.

2.5. *La volonté*

Selon la phénoménologie de Husserl, la conscience incarnée dans un corps matériel a également comme caractéristique essentielle de pouvoir se mettre en mouvement de manière spontanée et immédiate¹². Ainsi, on peut dire que la conscience est un « je veux » et un « je peux », elle a la potentialité de se mouvoir librement. Le sujet (la conscience phénoménale) est « un *ego* auquel appartient un corps en tant que champ de localisation de ses sensations ; il a la "faculté" ("je peux") de mouvoir librement ce corps et par conséquent les organes en lesquels ce corps s'articule et, par leur moyen, de percevoir un monde extérieur¹³ ». La liberté de la conscience se manifeste dans la spontanéité et l'immédiateté de l'acte.

3. *Les agents artificiels répondent-ils aux critères d'une conscience phénoménale ?*

Rappelons d'abord la thèse dont je tente de montrer l'inadmissibilité, c'est-à-dire celle des défenseurs de l'intelligence artificielle forte. Selon celle-ci, il serait envisageable de concevoir des agents artificiels dotés d'une conscience. En effet, certains experts en sciences cognitives - et plus particulièrement en neurobiologie computationnelle - croient « que le cerveau est un ordinateur numérique et que l'esprit conscient est un programme d'ordinateur [...] Ainsi construit, l'esprit est au cerveau ce que le logiciel (*software*) est au matériel (*hardware*)¹⁴ ». Le philosophe David Chalmers est l'un de ceux qui croient que les ambitions de l'intelligence artificielle forte sont raisonnables et que l'implémentation d'un bon calcul « suffit pour l'existence d'une expérience consciente aussi riche que la nôtre¹⁵ ». Si j'ai du mal à adhérer à la thèse selon laquelle une machine pourrait être dotée d'une conscience telle que nous la décrivons au début du texte - c'est-à-dire d'une expérience consciente qualitative - c'est parce que ces agents artificiels en question ne semblent pas pouvoir répondre aux

conditions de possibilités d'une conscience phénoménale que nous avons énumérées et développées dans les paragraphes précédents.

3.1. *Le corps matériel*

Comme nous l'avons vu, l'une des caractéristiques essentielles d'une conscience phénoménale est qu'elle soit incarnée. Or, aucun agent artificiel ne semble pouvoir répondre à ce critère. Par conséquent, tout porte à croire qu'aucun d'eux ne puisse prendre part à l'expérience consciente telle que nous la vivons. Bien qu'une intelligence artificielle soit installée dans un « corps physique » - ayant un support matériel - elle ne pourrait répondre au critère d'un corps biologique.

3.2. *La volonté*

Par ailleurs, même s'il était possible de reproduire artificiellement un support biologique, celui-ci serait privé d'un autre critère important, soit celui de la *volonté*. À ce sujet, dans les *Idées directrices* de Husserl, on peut lire que « les choses simplement matérielles ne sont susceptibles que de mouvement mécanique et la spontanéité de leur mouvement n'est que médiate¹⁶ ». Autrement dit, l'agent artificiel ne pourrait pas être doté d'une volonté libre, puisque les actes qu'elle commettrait seraient toujours médiatisés par un *programme* - nécessairement artificiel - qui traiterait l'information (*input* et *output*). Bien sûr, nous avons aussi notre cerveau qui agit comme système de traitement de l'information, mais la volonté implique un contenu de signification qui manque aux théories fonctionnalistes de l'esprit. Ainsi, la thèse de l'intelligence artificielle forte qui soutient que le fait « d'exécuter le bon programme *dans absolument n'importe quel matériel* est constitutif des états mentaux [...] le programme exécuté, par lui-même, garantit la vie mentale¹⁷ », ne pourrait voir le jour si l'on s'en tient à la condition d'un corps biologique muni d'une volonté libre. Néanmoins, allons plus loin encore afin de montrer le manque d'effectivité prévisible d'un tel projet.

3.3. La subjectivité et l'intentionnalité

Le nœud du problème est bien connu et c'est le suivant : comment peut-on envisager qu'un agent artificiel puisse parvenir à vivre une expérience qui inclut les traits de l'expérience phénoménologique ? Ce qui est le plus sujet à débat est ce qui concerne les qualités intrinsèques de nos vécus. La part subjective de notre expérience consciente est, en effet, l'élément le plus difficile à circonscrire. Si l'intelligence artificielle forte vise à réduire tout le fonctionnement du cerveau à un niveau de réalité physico-chimique, il est évident que leur entreprise est vouée à l'échec, car elle aura mis de côté une part essentielle de ce qu'est la conscience : le point de vue subjectif. Ainsi, en cherchant une objectivité plus grande par le processus de réduction, nous nous éloignerions de la véritable nature du phénomène plutôt que d'en avoir une vision plus claire¹⁸. À ce sujet, John Searle dira que « les états conscients n'existent que lorsqu'un sujet en fait l'expérience, et qu'ils n'existent que du point de vue à la première personne de ce sujet¹⁹ ». Les tenants de l'intelligence artificielle forte font l'erreur de croire qu'une simple manipulation de symboles pourrait expliquer l'ensemble de notre vie mentale. L'expérience phénoménale ne relève pas du domaine mathématique, mais du domaine de la *signification*. Or, l'erreur qui est faite est celle de confondre la syntaxe (la forme) et la sémantique (le contenu). Autrement dit, « l'esprit ne peut absolument pas se réduire à un programme d'ordinateur, car les symboles formels du programme de l'ordinateur ne suffisent pas en eux-mêmes à garantir la présence du contenu sémantique qui se produit dans les esprits réels²⁰ ». Comme il a été mentionné précédemment, la conscience est intentionnelle. Nous ne pouvons pas nous imaginer avoir des vécus sans contenu. Au contraire, notre expérience consciente a pour essence de viser des objets qui se phénoménalisent à nous par le moyen du sens. Or, le contenu de nos vécus relève de la signification.

3.4. L'auto-affection

De plus, si le caractère d'auto-affection est un trait essentiel d'une conscience phénoménale, je crois que le projet d'un agent artificiel conscient échouerait. Est-ce qu'un système computationnel pourrait véritablement être sujet et objet ? Autrement dit, est-il plausible qu'un tel système puisse parvenir à avoir une vision du monde comme étant *son* monde, selon *ses* objectivations, en plus d'être lui-même transformé par le monde ? Pourrait-il réellement être un sujet vivant dans un monde qui est le *sien* ? En revanche, puisque nous ne pouvons pas savoir quel effet cela fait d'être une intelligence artificielle, nous pouvons difficilement affirmer avec certitude l'impossibilité qu'elle se perçoive comme sujet - même si cette hypothèse me paraît peu probable.

3.5. La spatialité, la temporalité et la synthèse des vécus

Quoiqu'il en soit, il est important de mentionner que les caractéristiques d'une conscience phénoménale présentées précédemment étaient des critères essentiels pour qu'une telle conscience soit possible. Or, il suffirait d'un seul critère manquant chez l'agent artificiel pour que nous puissions dire que le projet de l'intelligence artificielle forte n'est pas soutenable. Comme nous le voyons, il semble que plusieurs conditions ne peuvent pas être remplies. Bien que j'admette que l'agent artificiel pourrait être compris comme ayant une sorte de *spatialité* et de *temporalité*, il est nécessaire de spécifier que cela ne serait que dans la mesure où la machine reçoit des données relatives au temps et à l'espace. Cette capacité aurait donc des traits similaires à celle qu'on attribue à la conscience phénoménale (se situer comme un « point zéro », avoir de la mémoire, etc.), mais elle s'en distinguerait par le fait qu'elle n'est pas le lieu d'apparition des phénomènes, qu'elle n'a pas d'expérience phénoménale et ce, parce qu'on peut difficilement lui attribuer une vie subjective. Le même phénomène se produit avec le critère de la synthèse des vécus. Nous pourrions penser que l'agent artificiel est apte à effectuer de telles synthèses, mais celles-ci concerneraient-elles des expériences vécues (comprises comme

Une conscience phénoménale est-elle possible chez des agents artificiels ?

des contenus de signification) ou des données (comprises comme des symboles formels) ?

En somme, on peut comprendre que l'agent artificiel ne répond pas à tous les critères d'une conscience phénoménale. En effet, il y a fort à parier qu'il ne pourrait être doté d'un corps matériel biologique en plus d'être muni d'une libre volonté. Par ailleurs, je crois que l'agent artificiel ne peut avoir accès à la part subjective de l'expérience phénoménale. Je pense que l'expérience de la machine se restreint à un programme délimité et qu'il lui est impossible de prendre part à une expérience vécue en tant que sujet. En ce sens, je ne crois pas que l'agent artificiel puisse avoir accès aux états intentionnels propres à l'expérience phénoménale. En effet, il est difficile d'imaginer que la machine ait accès au monde en tant qu'objet de signification. Je pense plutôt qu'on peut légitimement penser que son rapport au monde est uniquement mécanique, formel. À vrai dire, tout porte à croire que le monde se « donne » à l'agent artificiel par le biais de symboles vides - qui, ensemble, constitueraient une banque d'informations donnant à la machine l'allure de *comprendre* ce qui l'entoure - et non pas par le moyen du sens comme étant son monde. Or, dans la mesure où l'agent artificiel ne remplit pas l'ensemble des conditions de possibilité d'une conscience phénoménale, il nous est permis de douter, à juste titre, de la validité du projet de l'intelligence artificielle forte.

4. *Quelques problèmes en suspens*

Quoi qu'il en soit, le problème de la conscience phénoménale demeure un mystère difficile à élucider. Il est le point focal de toute la recherche en intelligence artificielle forte. Sans sa résolution, il semble que le projet d'une machine consciente pouvant expliquer la cognition humaine est difficile à imaginer. Une des questions qui demeure en suspens est celle de savoir : quelle science pourrait bien rendre compte de cette expérience phénoménale ? La méthode de recherche pouvant être fructueuse est toujours inconnue. En effet, les neurosciences peuvent expliquer certains éléments qui entrent en jeu dans la compréhension de la cognition humaine, mais comment peut-on expliquer de manière objective quelque chose qui concerne

le point de vue subjectif ? Autrement dit, dans l'optique où l'objet d'étude se vit à l'intérieur de soi, est-il prometteur de l'examiner à partir de l'extérieur ?

Finalement, au vu du développement technologique croissant, je crois qu'il nous manque certains éléments pour que l'on puisse trancher définitivement la question concernant la possibilité d'élaborer une machine consciente. À cet égard, établir les limites de la croissance technologique relèverait de l'expérience que nous en ferions dans l'avenir, donc relèverait de la spéculation. Autrement dit, rien n'empêche *logiquement* que la technologie se développe au-delà de ce que préalablement nous aurions établi comme étant une limite. Nous ne pourrions donc pas postuler *nécessairement* l'impossibilité de l'existence d'un agent artificiel conscient, mais cela ne nous empêche pas de défendre une telle thèse jusqu'à preuve du contraire. C'est pourquoi je ne m'empêche pas de soumettre mon hypothèse selon laquelle une telle idée est difficilement soutenable, étant donné que l'agent artificiel, tel qu'on peut l'imaginer, ne s'avère pas répondre à l'ensemble des caractéristiques essentielles d'une conscience phénoménale.

Toutefois, si l'on voulait arriver à une conclusion indubitable à ce sujet, il faudrait être en mesure de comprendre l'entièreté de la conscience phénoménale - ce qui relève d'un champ de la science qui demande encore d'importantes recherches. À ce sujet, l'hypothèse faisant consensus au sein des débats actuels en ce qui concerne les agents artificiels conscients est que nous manquerions de ressources pour prouver, hors de tout doute, la possibilité ou l'impossibilité d'un tel projet. Selon Paul Churchland, la conscience phénoménale nous paraît être inaccessible même si nous y avons accès en en faisant l'expérience à chaque instant : « nous échouons à reconnaître cette performance pour ce qu'elle est - un ballet computationnel subtil - parce que nous manquons de concepts et de ressources théoriques²¹ ». D'après lui, nous devons poursuivre les recherches en vue de développer l'intelligence artificielle, quel que soit le résultat effectif. Selon le postulat méthodologique de Churchland, les deux conséquences possibles seraient bénéfiques : « si on réussit à produire une machine vraiment intelligente, on peut mieux

connaître l'intelligence humaine en étudiant cette machine ; si, au contraire, on ne réussit pas, comme ce fut le cas jusqu'à maintenant, on peut approfondir notre connaissance de l'esprit humain en le contrastant relativement à la machine²² ». À cet effet, je seconde la proposition méthodologique de Churchland, mais je demeure perplexe, notamment quant à la nature des motivations qui animent la recherche en intelligence artificielle. Je crois qu'il serait nécessaire de se pencher sur la fin visée par de telles recherches et d'en examiner leur légitimité. Autrement dit, pour quelles raisons voudrions-nous créer une machine consciente ? Est-ce à des fins économiques ? Est-ce à des fins politiques ? Sommes-nous motivés par l'idée que l'intelligence artificielle apporte des solutions dans le domaine de la santé ? Est-ce que ces recherches cachent le désir de contrôler ou de déjouer la nature ? Je pense que les intentions derrière le projet de l'intelligence artificielle devraient être contestées, mais il va de soi que cela relève d'une tout autre discussion que celle dont on traite ici, soit de la *possibilité* même d'un tel projet.

En résumé, je remets en doute l'idée soutenue par les défenseurs de l'intelligence artificielle forte selon laquelle il serait possible de concevoir une machine réellement consciente qui serait dotée d'une vie mentale comme la nôtre. Si je conteste cette idée, c'est parce qu'il s'avère que l'agent artificiel ne remplit pas l'ensemble des critères d'une conscience phénoménale. En effet, certaines caractéristiques essentielles sont manquantes, notamment celle d'avoir comme support un corps biologique muni d'une spontanéité et d'une volonté libre. Néanmoins, ce qui rend l'idée d'un agent artificiel conscient encore plus improbable, c'est qu'elle ne semble pas prendre en compte la part proprement subjective de l'expérience consciente. En effet, en réduisant la conscience phénoménale au cerveau et à ses processus neuronaux, les tenants de l'intelligence artificielle forte passent à côté d'un élément essentiel à la possibilité même d'une conscience phénoménale et donc également de la validité de leur projet. Cet élément est la part subjective de l'expérience, c'est-à-dire le point de vue du sujet auquel les choses se phénoménalisent par le moyen du sens. Si l'on dit que la conscience phénoménale est intentionnelle, c'est parce qu'elle accède au monde par le biais de ses vécus et ces vécus

de conscience ne sont réductibles à aucune explication objective. Il va de soi que ma réflexion portant sur l'idée d'une conscience phénoménale chez les agents artificiels porte davantage sur ce qui est accessible, préférant mettre entre parenthèses le champ spéculatif, c'est-à-dire l'ensemble des connaissances qui sont actuellement hors de notre portée et dont la validité n'est que probable. En effet, il reste plusieurs questions en suspens et nous pouvons prévoir une croissance technoscientifique dont l'étendue nous est présentement inconnue. Toutefois, les agents artificiels ne semblent pas, encore à ce jour, pouvoir répondre aux conditions de possibilité d'une conscience phénoménale et c'est pourquoi le projet de l'intelligence artificielle forte me paraît peu réalisable.

-
1. Il va sans dire que plusieurs ne considèrent pas nécessairement ces différences de niveaux comme étant inconciliables. Évidemment, il aurait lieu de nuancer davantage en évitant de catégoriser ainsi, mais afin d'alléger le texte, je m'en tiendrai à cette distinction.
 2. Les qualia sont le contenu subjectif d'une expérience. Ils peuvent être une sensation physique, une émotion, une simple perception, etc.
 3. Les propriétés ayant une réalité empirique, matérielle.
 4. Edmund Husserl, *Idées directrices pour une phénoménologie et une philosophie phénoménologiques pures, Livre Second, dans Recherches phénoménologiques pour la constitution*, trad. Éliane Escoubas, Paris, PUF, 1982, p. 208.
 5. La corporéité de la conscience sera développée également dans les sections ultérieures.
 6. John R. Searle, *Le mystère de la conscience*, Paris, Éditions Odile Jacob, 1999, p. 30.
 7. Thomas Nagel, *Questions mortelles*, trad. P. Engel, Paris, PUF, 1983, p. 399.
 8. Edmund Husserl, *Recherches logiques, Tome 2, Recherches pour la phénoménologie et la théorie de la connaissance* [1901], deuxième partie : Recherche V, trad. H. Élie, A. L. Kelkel et R. Schérer, Paris, PUF, 1969, p. 146-147
 9. Edmund Husserl, *Méditations cartésiennes. Introduction à la phénoménologie* [1947], trad. G. Peiffer et E. Levinas, Paris, Vrin, 2014, p. 75.

Une conscience phénoménale est-elle possible chez des agents artificiels ?

10. Edmund Husserl, *Idées directrices pour une phénoménologie et une philosophie phénoménologiques pures*, *op. cit.*, p. 262
11. *Ibid.*, p. 265.
12. *Ibid.*, p. 215.
13. *Ibid.*, p. 215-216.
14. John R. Searle, *op. cit.*, p. 21.
15. David Chalmers, *L'esprit conscient. À la recherche d'une théorie fondamentale*, Paris, Les Éditions d'Ithaque, 2010, p. 431.
16. Edmund Husserl, *op. cit.*, p. 215.
17. John R. Searle, *op. cit.*, p. 26.
18. Thomas Nagel, *op. cit.*, p. 400.
19. John R. Searle, *op. cit.*, p. 128.
20. *Ibid.*, p. 23.
21. Paul Churchland, *Le cerveau. Moteur de la raison, siège de l'âme*, Paris, Bruxelles, De Boeck Université, 1999, p. 251.
22. Serge Robert, « Réflexion épistémologique sur l'intelligence artificielle et les sciences cognitives : à quelles conditions une machine pourrait-elle connaître ? », dans *Philosophie et sciences : du concept au réel*, Vol. 2, no°2, printemps 1992, p. 173.