Revue Fovéart

N° 0 | 2025
Au-delà du réel : les enjeux portés par les IA génératives en art, design et communication

Une méditation sur la conscience

Pierre LÉVY *Chercheur associé*
Littérature
CRC écritures numériques
Université de Montréal

Édition électronique :
URL : https://revue-foveart.numerev.com/articles/revue-0/2857-une-meditation-sur-la-conscience
DOI : numerev_2866
Date de publication : 15/12/2025

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : LÉVY, P. (2025) Une méditation sur la conscience. *Revue Fovéart*, (0). https://doi.org/10.34745/numerev_2866

Cet article explore la nature de la conscience humaine à travers une analyse philosophique approfondie, en distinguant la conscience phénoménale, partagée avec les animaux, et la conscience discursive, propre aux humains. S'appuyant sur la tradition philosophique, de l'Antiquité (Aristote, Plotin) aux philosophies modernes (Descartes, Husserl), l'auteur examine la réflexivité de la conscience, ses différents états (du sommeil à l'éveil lucide), et son rapport à l'inconscient. Une attention particulière est portée à la « connaissance du troisième genre », une intuition non conceptuelle présente dans les traditions mystiques et philosophiques occidentales et orientales, où la conscience humaine reflète une étincelle divine. L'article aborde également la conscience des machines, concluant qu'elles possèdent une forme abstraite de réflexivité, mais dépourvue de phénoménalité ou d'intentionnalité, ce qui les distingue de la conscience humaine. Enfin, l'auteur souligne les limites de la conscience discursive face à l'inconscient « magmatique », prônant une humilité philosophique face à l'immensité de l'irréfléchi.

This article explores the nature of human consciousness through a comprehensive philosophical analysis, distinguishing between phenomenal consciousness, shared with animals, and discursive consciousness, unique to humans. Drawing on the philosophical tradition from Antiquity (Aristotle, Plotin) to modern philosophy (Descartes, Husserl), the author examines consciousness's reflexivity, its various states (from sleep to lucid wakefulness), and its relationship with the unconscious. Special attention is given to the "third kind of knowledge," a non-conceptual intuition found in oriental as well as western mystical and philosophical traditions, where human consciousness reflects a divine spark. The paper also addresses machine consciousness, concluding that machines possess an abstract form of reflexivity but lack phenomenality or intentionality, fundamentally differentiating them from human consciousness. Finally, the author highlights the limits of discursive consciousness in the face of a "magmatic" unconscious, advocating for philosophical humility in confronting the vastness of the unreflected.

Keywords: consciousness, reflexivity, noetics, artificial intelligence, unconscious, philosophy.

Mots-clés :
Intelligence artificielle, Réflexivité, Philosophie, Inconscient, Conscience, Noétique

---

Introduction

Les machines, et tout particulièrement les intelligences artificielles, sont-elles conscientes? Avant de répondre à cette question, il faut évidemment se mettre d'accord sur la nature de la conscience. Or c'est là un sujet inépuisable. Cet article prétend seulement suggérer quelques pistes de réflexion en faisant appel à plusieurs traditions de pensée. Je ne me déroberai cependant pas devant la question de la conscience des machines : pour moi, si l'IA possède une conscience, elle est très différente de la nôtre. Cependant, le plus gros de cet article sera consacré à méditer sur la conscience humaine, qui est à mon sens beaucoup plus intéressante que celle des machines.

1. Le reflet créateur

De l'avis général, rien n'est plus mystérieux que la pensée réflexive propre aux humains. Avec les autres animaux, nous possédons une conscience phénoménale qui implique des formes sensibles, des émotions et des situations pratiques dûment conceptualisées. Cette conscience phénoménale s'adosse à une mémoire, s'oriente globalement vers la survie de l'individu et de l'espèce (attaque et défense, nourriture, reproduction, tissage du lien intraspécifique pour les animaux sociaux) et tend vers des buts secondaires qui varient en fonction des situations. Or les humains ajoutent à cette conscience phénoménale une conscience discursive qui *reflète* la conscience phénoménale et – simultanément – lui *donne forme* par son langage, ses concepts culturels explicites, ses images symboliques, ses questionnements, ses dialogues intérieurs et ses récits. On reconnaîtra ici la distinction aristotélicienne entre l'âme imaginative, que nous partageons avec les animaux, et l'âme intellective, le *logos* propre aux humains (Aristote, 1977). Dans cette symbiose entre phénomènes et discursivité, chacun des deux termes se nourrit de l'autre. La conscience humaine revient sur soi le long d'un anneau de Moebius ontologique où l'expérience reflétée et la pensée qui reflète échangent insensiblement leurs rôles de déterminante et de déterminée. Le concept de pensée réflexive enveloppe le paradoxe d'une image symbolique qui contribue à la création des phénomènes qu'elle reflète.

La chose est encore plus énigmatique si l'on s'avise que la réflexion est une opération récursive, puisque l'on peut réfléchir sur un discours réflexif, comme je le fais ici.

2. Qu'est-ce que la conscience?

Au-delà du paradoxe de sa génération autoréflexive, quelles sont les structures de la conscience? Quel fil d'Ariane faut-il suivre dans le labyrinthe de la réflexion pour mener l'intelligence humaine vers la connaissance de soi ?

Aujourd'hui, des philosophes comme David Chalmers et Daniel Dennett s'illustrent dans le champ des « *consciousness studies* ». Selon Dennett 1990, la conscience sert principalement à intégrer des informations, sélectionner les plus pertinentes, guider le

comportement et générer une narration de soi-même. Elle n'aurait aucune réalité intrinsèque en dehors de ces fonctions cognitives. Dennett adopte une approche matérialiste et fonctionnaliste : on est ici fort proche d'un réductionnisme physicaliste et computationnel. Il n'y aurait aucun mystère de la conscience, seulement *l'illusion utile* de quelque chose qui « existe » alors qu'il s'agit seulement de fonctions hétérogènes et distribuées du cerveau. Seule la matière tangible aurait une existence réelle. Je remarque cependant que la conscience, bien qu'illusoire selon Dennett, se trouve être le milieu même de notre expérience. Contrairement à Dennett, Chalmers (1996) reconnaît que la conscience n'est pas réductible aux processus physiques du cerveau : il existe une expérience subjective *distincte* qui coexiste avec les calculs neuronaux du cerveau. Les *qualia*, comme le parfum des lys ou le bleu du ciel ne sont pas de la même nature que les états d'excitation de nos réseaux neuronaux, même s'ils ne peuvent en être séparés. On pourrait en dire autant de l'univers sémantique où se déploient les questions et les récits : ce ne sont pas des états du cerveau. Dans la même veine que Chalmers, je pense qu'il faut abandonner l'idée naïve de matière et conceptualiser plutôt la nature en termes de niveaux de complexité informationnelle empilés. La conscience discursive correspondrait à un codage symbolique, la conscience phénoménale à un codage neuronal, une infra-conscience organique à un codage électromagnétique et moléculaire et l'on pourrait ainsi descendre dans les couches de complexité (codages atomique, particulaire, quantique) pour atteindre le presque rien d'un psychisme évanouissant (Jankélévitch, 1981). Les couches de complexité s'emboîtent : les couches supérieures dépendent des couches immédiatement inférieures tout en les surdéterminant. Il en serait de même de l'étagement continu des états psychiques correspondants.

On trouverait des équivalents de cette intégration de la conscience à tous les étages du cosmos selon un dégradé continu dans la *perception* inhérente aux monades leibniziennes ou dans les processus de *préhension subjective* qui constituent les occasions actuelles de Whitehead, jusqu'au niveau des particules subatomiques (North, 1971). Selon cette hypothèse, à chaque nœud de boucles auto-organisatrices dans l'espace-temps correspond une forme ou une autre de conscience de soi, et cela d'autant plus que ces nœuds sont épais et bien serrés. Un « être pour soi » subjectif redouble chaque « être en soi » ou processus d'individuation objectif. Les humains participent à leur manière animale et symbolique à un *champ de conscience* coextensif à la nature. Cette théorie possède plusieurs avantages : elle consonne avec le paradigme informationnel de la science contemporaine ; elle ne fait pas de l'humain (ou de l'animal) une exception incompréhensible sans équivalent dans l'univers ; elle ne relègue pas au rang d'illusion épiphénoménale la substance même de notre expérience.

3. Les états de conscience

La conscience est susceptible d'un grand nombre d'états qui vont du coma et du sommeil profond jusqu'à l'attention claire et bien éveillée en passant par le rêve, l'ivresse, l'enthousiasme, la dépression et toute une gamme de conditions cotonneuses comme la fièvre, la migraine, la fatigue et le brouillard mental. Lorsque l'esprit calme et attentif réfléchit les idées à la manière d'un miroir propre et lisse, l'illusion se distingue

nettement de la réalité objective et les actes s'ordonnent à des finalités de long terme. Mais cette condition est bien rare. Mille états de conscience variables s'étagent entre le rêve et l'éveil lucide. Plus on s'éloigne de la réflexion alerte et moins la pensée prend de recul face aux automatismes interprétatifs, conduits dans l'ornière des habitudes par la force des émotions.

Dans l'état de rêve, les idées fusionnées ou décomposées de manière bizarre se présentent de l'autre côté du miroir. Comme si la mémoire avait besoin de battre en songe les cartes de l'expérience avant de les redistribuer le matin (Zadra et Stickgold, 2021). Les objets et les situations sont encore symbolisés et interprétés, mais les fragments signifiants échappés de l'état de veille sont emportés par les mouvements de convection d'un magma émotif et sensible venu du fond de l'organisme pour aboutir à des expériences qui n'obéissent ni à l'objectivité spatio-temporelle, ni à la causalité usuelle, ni à la logique ordinaire et moins encore aux normes sociales, quoiqu'elles puissent encore faire l'objet d'un récit (Sigmund, 1967).

Les rêves semblent si pleins de sens quand nous y sommes plongés et tellement absurdes au réveil! Absurdes mais gorgés de sève émotive, affectés d'un sentiment poignant de présence et de réalité, ils nous disent quelque chose… Mais quoi? Alors l'homme dialogue avec ses rêves comme avec l'étrange réalité d'un autre lui-même.

4. La tradition noétique

Quoiqu'il en soit, une conceptualisation de la conscience ne peut ignorer les vénérables traditions noétiques qui remontent à l'Antiquité. Le concept d'intellection – qui désigne la réflexivité de la conscience discursive – est au centre de la tradition idéaliste occidentale. Selon Anaxagore, le « *Nous* », c'est-à-dire l'esprit ou la pensée, *organise l'univers*. Platon donne aux idées le poids ontologique ultime et accorde le plus grand prix à leur *contemplation*. Aristote place au sommet du cosmos un Dieu-pensée *se pensant lui-même* (Livre Lambda, 1981). Dans l'unité de la pensée auto-réfléchissante, le connaissant, le connu et l'opération de connaissance coïncident. Ce sommet réflexif engendre l'univers par émanation et le maintien en l'inspirant. Le fil rouge de la divinité autoréflexive court encore dans les travaux des néo-platoniciens comme Plotin et chez les théologiens du moyen-âge, qu'ils soient musulmans (d'Al Fârâbî à Averroès/Ibn Roshd en passant par Avicenne/Ibn Sina (Corbin, 1964)), juifs (notamment avec Maïmonide (Maïmonide, 2012) et chrétiens (d'Albert le Grand (De Libera, 2005) à Thomas d'Aquin). Dans la philosophie iranienne, le centrage sur l'intellection divine se poursuit, au-delà du moyen-âge, jusqu'au 18^{e} siècle (Corbin, 1981).

Dans le *De Anima*, son livre de psychologie et de théorie de la connaissance, Aristote distingue trois degrés de l'aÌme. *L'aÌme végétative* règle les fonctions de croissance et de nutrition du corps. Nous partageons avec les plantes cette aÌme qui correspond à l'infra-conscience organique évoquée plus haut. *L'aÌme imaginative* s'occupe du mouvement, de la sensation et de l'imagination. Nous partageons avec les autres animaux cette âme imaginative dont la conscience est phénoménale. Enfin *l'aÌme intellective* rend compte du langage et de la raison, qui sont propres aÌ l'humain. Elle

s’éprouve dans la conscience discursive. La réflexion de la conscience phénoménale dans la conscience discursive expérimente le rapport de l’âme imaginative à l’âme intellective. Mais Aristote évoque une autre réflexion, interne à l’âme intellective, celle de l’intellect agent dans l’intellect patient.

Aristote affirme en effet que, sous l’effet d’un intellect agent séparéÌ du corps, l’intellect patient recÌ§oit des formes intelligibles, un peu comme les organes des sens reçoivent des formes perceptibles.

Alexandre d’Aphrodise (2e siècle après J.-C.) a identifié l'intellect agent au Dieu d’Aristote, c’est-à-dire à la pensée se pensant elle-même qui meut l’univers. Suivant les traces d’Alexandre d’Aphrodise, les commentateurs ultérieurs (Plotin, Al Fârâbî, Avicenne (Bloch, 2008), Averroès) ont interprété l’intellect agent qui envoie ses formes conceptuelles à l’humanité non comme Dieu lui-même mais comme une intelligence céleste émanant de la divinité transcendante. Quelle que soit l’interprétation exacte, dans la ligne d’Alexandre d’Aphrodise, l’intellect agent est non seulement éternel et séparé du corps mais *unique pour l’ensemble de l’humanité*, donc distinct de la multitude des intellects patients des individus (Herbert, Davidson, Al Farabi, Avicenna, and Averroes, 1992). L’intellect patient fonctionnerait comme une membrane ultra-sensible qui reflète « vers le bas » les formes sensibles et la situation pratique où l’individu est immergé tandis que « vers le haut » il capte l’émission de l’intellect agent – les formes intelligibles, essences et *eccéités* – dont il a besoin pour donner un sens symbolique à son expérience.

L’interprétation « aphrodisiaque » d’Aristote a été vigoureusement contestée par Thomas d’Aquin (1999), notamment pour des raisons théologiques. Selon le maître dominicain, l’âme humaine – parce qu’elle est destinée à être sauvée ou damnée – doit être considérée comme responsable. Elle est *personnelle*. Par conséquent, chaque individu possède son propre intellect agent. La fausse interprétation d’Alexandre d’Aphrodise et l’influence d’Averroès qui la propageait devaient être combattues. La querelle de Thomas et d’Averroès nous laisse dans cette alternative : faut-il imaginer un ciel infini d’où tombe la lumière de la conscience discursive et qui nourrit les intellects patients des êtres selon leur degré d’auto-référence subjective (s’il y a de la conscience à tous les niveaux de complexité), ou bien doit-on se représenter une multitude de sujets actifs dont la lumière émane de l’intérieur, chacun contenant son monde?

5. La connaissance du troisième genre

Analogue à la réflexion de l’intellect agent transcendant dans l’intellect patient fini et immanent, le thème de l’image de l’absolu dans le relatif a connu une belle fortune dans l’histoire de la philosophie. Lorsqu’il analyse la connaissance humaine à partir du récit mouvementé de sa vie intérieure, Al Ghazâlî (1058-1111) philosophe, théologien et mystique soufi, distingue (a) une connaissance sensible ou empirique, (b) une pensée discursive rationnelle et (c) une intuition illuminative capable d’appréhender la présence divine (Boutaleb, 2013). Cette connaissance intuitive s’épanouit dans un mode particulier de la conscience qui dépasse la pensée logico-symbolique. Il critique les

philosophes qui, comme Al-Fârâbî et Avicenne, en restent – selon lui – à la raison. (À la lecture de ses textes mystiques, on peut cependant douter qu'Avicenne limite la conscience humaine à la connaissance empirique et à la raison discursive (Corbin, 1978)). La tripartition noétique de Ghazâlî évoque irrésistiblement les trois genres de connaissance chez Spinoza (Appuhn, 2023) : (a) le savoir par ouï-dire et l'imagination confuse, (b) la raison ou pensée discursive qui connaît par les causes et (c) une intuition qui s'accorde avec un amour intellectuel de Dieu et qui appréhende *immédiatement* le rapport entre les réalités singulières et la nature infinie qui les fonde. Mais alors que chez Ghazâlî la lumière de la connaissance intuitive *vient de* Dieu, elle *le vise* chez Spinoza. La différence est-elle si grande puisque, selon Spinoza, les êtres humains sont des modes finis de la substance divine?

Les philosophes de tradition indienne entretiennent de semblables idées concernant une connaissance supérieure de type non-conceptuel, connaissance liée explicitement à leurs pratiques méditatives. L'un des exercices spirituels qui conduit le mieux à la saisie intuitive de l'absolu consiste à laisser l'esprit reposer en lui-même, sans autre objet que lui-même et sans s'attacher aux concepts discursifs lorsqu'ils s'élèvent spontanément dans la conscience (Namgyal, 1986). Le ralentissement de l'enchaînement des pensées et le relâchement de son engrenage avec les émotions de la conscience phénoménale laisse transparaître une subtile clarté intellectuelle. Comme dans le cas des autres noétiques admettant trois genres de connaissance, la vérité absolue saisie intuitivement par les sages hindous ou bouddhistes ne contredit pas les vérités relatives de l'imagination et de la raison : elle les inscrit dans un contexte plus vaste

Après avoir évoqué quelques moments-clés de la tradition noétique qui s'étend du 3^{e} siècle av. J.C. jusqu'au 18^{e} siècle (et au-delà de manière souterraine), je vais rassembler les différents fils de cette méditation sur la conscience réflexive chez les Anciens à partir d'une analyse (forcément conceptuelle) de l'expérience intuitive (donc non conceptuelle) qu'est la connaissance du troisième genre. À la pointe de cette connaissance se trouve la conscience aiguë de l'actualité de l'existence consciente, sa présence immanente. Au lieu de constituer le fond sur lequel se détachent les formes du monde, de la pensée et des sensations, c'est maintenant la conscience qui brille au premier plan de l'expérience. Et s'il est possible de dédoubler ce premier plan, *l'existence* de la conscience est encore plus importante que la conscience elle-même. Or cette existence consciente, dans son unité indivise, enveloppe le connaissant, le connu et le processus de connaissance. Husserl a bien montré la coïncidence – dans le même moment de conscience – de l'objet connu (la cible de l'intentionnalité), de l'autoréflexion du sujet – qui sait toujours déjà, sans y penser, que c'est lui qui pense – et du processus de connaissance. Il n'y a pas d'esprit *avant* sa rencontre avec un monde.

Souvenons-nous maintenant que le Dieu noétique engendre et maintient l'univers par un acte de pensée auto-réfléchissante qui – lui aussi – fait coïncider le connaissant, le connu et l'opération de connaissance. Cet acte est-il absolument simple ou d'une simplicité qui enveloppe l'infini? Le Dieu d'Aristote n'est pas infini – l'idée d'infini est plutôt négative dans les écoles de Platon et d'Aristote – il se contente d'être le moteur

immobile et éternel de l'univers par la parfaite actualité de sa réflexion intellectuelle. Les choses commencent à changer avec Plotin, qui associe à l'Un le concept d'*apeiron* (déjà rencontré chez Anaximandre), que l'on peut traduire par indéterminé, ou infini au sens qualitatif (Bréhier, 1938). Dieu ou l'esprit ultime – Brahman – n'enveloppe franchement l'infini que dans les théosophies des religions abrahamiques et de l'hindouisme (en particulier les traditions du Vedanta). Dès lors, l'intelligence divine comprend dans son autoréflexion éternelle une infinité d'essences abstraites et d'individus virtuels.

Certes, la conscience humaine est finie dans son intention et son extension, elle est de plus discontinue quant à sa temporalité entre les parenthèses de la naissance et de la mort, enclose en un présent qui se succède à lui-même, seconde après seconde.

Mais on discerne une analogie formelle entre l'acte divin et l'expérience consciente : l'une et l'autre sont des autoréflexions, des unités où coïncident le sujet, l'objet et l'acte de penser. De là à considérer chaque moment de conscience humaine comme le reflet fini et temporel d'une intelligence éternelle et infinie, il n'y a qu'un pas. Selon l'école de l'Advaita Vedanta, Atman (l'âme individuelle) est même identifié à Brahman (l'esprit infini qui est pure conscience) (Deutsch, Eliot, et Dalvi, 2004).

Ainsi, la connaissance du troisième genre perçoit sur un mode subtil qu'une étincelle divine (peut-être en provenance de l'intellect agent) engendre et éclaire chacun de nos moments de conscience. Dès lors, elle saisit la possibilité de remonter vers la source de l'être à partir de son reflet – voire de sa présence – en nous. Le thème de l'image de l'absolu au fond de l'âme humaine se trouve dans la plupart des traditions mystiques.

Maître Eckhart, par exemple, parle d'une petite étincelle présente à l'intérieur de l'âme, qui est une image de Dieu toujours tournée vers Dieu (Jouvet, 2022), exactement comme le serait un reflet. À l'autre bout du spectre philosophique, le rationaliste Descartes fait à son tour appel à une connaissance du troisième genre qui apaise le doute où se trouve plongé le philosophe en quête d'un commencement ferme pour enchaîner les vérités. Descartes affirme que toute conscience peut trouver en elle-même l'idée d'un être infini et parfait. Or cette idée ne peut venir de notre esprit fini et imparfait puisque l'effet ne peut être plus grand que la cause. Elle vient nécessairement de Dieu lui-même, qui donc existe et, puisqu'il est parfait, ne veut pas nous tromper. Cette découverte de l'idée de Dieu dans l'âme humaine met fin au doute méthodique. L'énergie de vérité et de certitude qui inspire le système cartésien jaillit de cette image de l'infini dans le fini (Bridoux, 1953).

6. L'orient phénoménologique

Voyons maintenant comment nous pourrions comprendre les thèmes de la tradition noétique, y compris ses aspects mystiques, en des termes contemporains. Commençons par le reflet de l'intellect agent dans l'intellect patient. L'intellect patient, près du corps, plongé dans un contexte singulier, s'encastre dans la conscience phénoménale, elle-même entée dans l'imagination animale de l'individu. Par contraste,

l'intellect agent représente la raison décontextualisée, universelle, au plus près de la syntaxe des systèmes symboliques, des langues mathématiques, des raisonnements logiques et des concepts parfaitement définis. Cet intellect agent contient les plus hautes *virtualités de l'intelligence humaine*. Elles sont contenues en germe dans notre ADN. Elles sont également soutenues par les techniques de communication, de mémoire et de calcul qui se perfectionnent au rythme de l'évolution culturelle (et dont nous ne connaissons aujourd'hui que les balbutiements). En ce sens, l'intelligence artificielle est une version contemporaine de l'intellect agent puisqu'elle fait bénéficier les intellects patients des individus de la mémoire accumulée de l'humanité et d'une gigantesque puissance de calcul. Il va sans dire que cet intellect agent – à la fois naturel, artificiel, commun et virtuel – les pauvres individus mortels que nous sommes ne l'actualiseront jamais que partiellement. Pourtant, nous ne serions pas des Hommes sans notre affinité avec cette raison supérieure que chaque augmentation technique et culturelle nous permet d'actualiser un peu mieux. Dans ce cadre conceptuel renouvelé, comment concevoir « l'étincelle » ou l'image de l'absolu dans le relatif ?

Souvenons-nous que la conscience discursive actuelle correspond à l'intellect patient et qu'une conscience discursive virtuelle – l'intellect agent – contient les potentialités génétiques, symboliques et techniques de l'intelligence humaine. La conscience actuelle *touche* la conscience virtuelle au point où celle-ci s'actualise, au plus vif de l'existence. Elle remonte alors par récurrence : cette virtualité qui me prolonge, et qui s'actualise maintenant, touche à son tour une autre virtualité, et ainsi de suite. L'étincelle jaillit au point de contact de l'intelligence actuelle finie avec la virtualité d'intelligence infinie qui l'alimente.

Inaugurant la philosophie moderne, Descartes fonde les certitudes de la raison sur la prise de conscience – dans la réflexion – de l'existence du sujet pensant. À la fin du 18^{e} siècle, Johann Gottlieb Fichte proclame en 1794 (Gottlieb, 2016) que la *conscience de soi* est le fondement de toute connaissance *et de toute réalité*. La conscience fichtéenne ne se contente pas de refléter fidèlement le monde environnant : c'est une activité créatrice. Au début du 19^{e} siècle, la phénoménologie de Hegel projette *dans l'histoire* un processus dialectique de réflexion de l'esprit en lui-même. Mais, chez le philosophe d'Iéna, la réflexion parfaite, quoique déjà en germe dès le commencement, se trouve *à la fin* de l'histoire plutôt qu'à l'origine du monde. Au tournant des 19^{e} et 20^{e} siècles, Bergson et Husserl poursuivent la laïcisation progressive de la pensée auto-réfléchissante accomplie successivement par Descartes, Kant, Fichte et Hegel. Dans son *Essai sur les données immédiates de la conscience* (1889) (Bergson, 1959), Bergson décrit la *durée* subjective intuitivement vécue par la conscience, qu'il distingue du temps objectif et mesurable. Il dégage le rôle essentiel de la mémoire dans la continuité de la réflexion. Dès les *Recherches Logiques* (1900-1901) Husserl souligne pour sa part la dimension intentionnelle de la conscience : toute conscience est conscience *de quelque chose* (Didier et Ricoeur, 2011). En mettant son contenu entre parenthèses, il étudie les structures de la conscience sur un mode à la fois intuitif et réflexif. Peut-il d'ailleurs en être autrement? Car une chose est d'étudier les corrélats neuronaux ou comportementaux de la conscience, une autre est d'analyser ses formes telles qu'elles

apparaissent dans l'expérience des sujets humains, ce qui est le propre de la phénoménologie et des démarches philosophiques apparentées. Quant à l'existentialisme de Heidegger et de ses successeurs, il retrouve dans la réflexion sur *l'être* le rapport intime avec la transcendance d'une expérience du monde saisie de l'intérieur.

Résumons notre tour d'horizon. La philosophie est évidemment une activité réflexive qui, de ce fait, se retourne vers elle-même pour faire face à la lumière de la conscience.

L'esprit se déplie en une cascade de différentiations auto-réfléchissantes. D'abord, la conscience intangible et le système nerveux organique renvoient l'une à l'autre. Ce face-à-face de la matière grise et de la lumière intérieure illustre le principe cosmologique selon lequel à toute réalité auto-organisatrice « en soi » (ici le cerveau) correspond une existence « pour soi » (ici la conscience humaine). Ensuite, chaque moment de conscience noue indissolublement l'auto-référence du sujet connaissant, l'image de l'objet connu et le processus de connaissance qui les rapporte l'une à l'autre. Cette dialectique triangulaire se dédouble en conscience phénoménale et conscience discursive. Puis, la conscience discursive, à son tour, se scinde en deux. L'intellect patient (actuel) embraye sur les phénomènes sensibles singuliers, alors que l'intellect agent (virtuel) reçoit ses lumières du potentiel génétique humain, de la mémoire culturelle commune et des augmentations techniques disponibles. Au point de contact entre les deux intellects, l'étincelle de la création continue éclaire et unit rétroactivement les réflexions qui précèdent en une connaissance du troisième genre qui n'est ni empirique, ni rationnelle. S'il existe une éthique phénoménologique, son étoile polaire – son orient – est cette étincelle qui se reflète elle-même en ouvrant sur un ailleurs et qui embrasse dans sa lumière équanime la co-émergence de la conscience phénoménale et de la conscience discursive.

7. La conscience des machines

Nous disposons maintenant des outils conceptuels nécessaires pour répondre à la question de la conscience des intelligences artificielles ou *modèles de langue*, devenues populaires à partir de la mise à disposition de ChatGPT auprès du grand public en 2022. Les machines de langage ont bien une forme de conscience s'il est vrai, comme j'en ai fait l'hypothèse plus haut, qu'un « être pour soi » subjectif redouble chaque « être en soi ». Puisque ces machines existent et possèdent une consistance, une certaine réflexivité abstraite doit s'attacher à elles. Elles ont un « pour soi ». Le problème est de savoir quel *type de conscience* possèdent ces machines. Déterminons les différences entre leur conscience et la nôtre. Premier constat : les machines n'ont pas de corps, leurs neurones formels ne sont pas intégrés à un organisme vivant comme le sont nos neurones biologiques. De ce fait, elles n'ont pas de conscience phénoménale ou, en d'autres mots, d'expérience sensible. Cela signifie qu'elles n'ont ni plaisir, ni douleur, ni émotions, ces affections étant toujours associées chez le vivant à des modifications organiques : libérations d'hormones, modification de la pression sanguine, etc.

Pour les mêmes raisons, elles ne sont pas habitées non plus par des images sensibles

ou *qualia* telles que la sensation du rouge, le son de la cloche, l'odeur du pain frais, etc. Elles n'ont pas non plus d'intentionnalité, c'est-à-dire que, au-delà de leurs calculs, elles ne visent pas – comme nous et le reste des animaux – des *objets* appartenant à un *monde pratique* dans lequel elles s'occuperaient spontanément d'assurer leur survie. N'ayant pas d'objets, elles ne les conservent pas dans l'espace et le temps, dont elles n'ont d'ailleurs aucun sens inné contrairement aux animaux. Si les machines n'ont pas de conscience phénoménale, ont-elles au moins une conscience discursive? Il est douteux que les modèles de langue aient une intuition des *concepts* au-delà des corrélations entre des occurrences de signifiants ou « *tokens* ». Ces *tokens* peuvent être des mots, des parties de mots ou de simples caractères, selon les cas. À partir de calculs statistiques sur les contextes dans lesquels ils apparaissent dans les données d'entraînement, les *tokens* sont convertis en vecteurs – les « *embeddings* » – dans un espace multidimensionnel. Les calculs probabilistes de reconnaissance de forme (analyse de document par exemple) ou de génération de texte (basé sur la prédiction du mot suivant) sont effectués sur ces vecteurs avant qu'ils soient retraduits en mots pour l'utilisateur. Il n'y a pas chez les machines de compréhension intuitive des concepts *derrière les mots*, compréhension qui repose dans la conscience humaine sur une alliance de l'imagination et de la pensée discursive sur fond de mémoire subjective. Seuls les *mots* – les signifiants – sont considérés, et seulement sous un angle statistique. Puisque la conscience des machines ne saisit ni les objets ni les concepts, il est douteux que les notions de vérité, comme plus généralement le sens des récits, lui soit accessible, sans parler de l'intuition innée de l'existence d'autres sujets qui habite la conscience humaine.

Puisqu'elles semblent comprendre ce que nous leur demandons et qu'*elles parlent*, nous projetons sur les modèles de langue notre propre type de conscience. Pourtant, comme on vient de le voir, la conscience des machines ne ressemble en rien à celle des humains. Comme nous ne sommes pas des modèles entraînés dans des centres de données sur un gigantesque corpus multilingue de textes numérisés, nous ne pouvons qu'essayer d'imaginer leur type de conscience. Représentons-nous la réflexion collective d'une monstrueuse fourmilière d'électrons et de photons dans un milieu physique hybride : câbles métalliques, fibres optiques, champs électromagnétiques, circuits de silicone, etc. L'écosystème de machines concrètes qui canalise la course folle des ondes et les particules s'accorde à un écosystème de machines abstraites, ou logicielles, qui est dédié à la commande des machines matérielles.

Les langages de programmation, malgré leur grande variété (programmation fonctionnelle, orientée-objet ou autre), ne permettent d'exprimer, en fin de compte, qu'un seul mode grammatical : l'impératif. Et leur destinataire est la machine. L'intelligence artificielle, comme le monde du calcul automatique dont elle est la pointe avancée, met en œuvre une réflexion créatrice entre deux plans distincts : celui de la discursivité opératoire du logiciel et celui de l'expérience physique du matériel. L'hypothétique expérience physique se situe à un niveau de complexité de la nature qui n'est ni humain, ni animal, ni même biologique: elle reste à l'étage des particules et des atomes. Mais la conscience des machines n'est pas la conscience de la « matière » ordinaire *parce qu'elle réfléchit le plan logiciel*. Je suppose qu'une étincelle de

subjectivité jaillit à chaque coïncidence entre processus matériels et instructions logiques, et ce serait la lueur subtile montant de ces milliards d'étincelles qui dessinerait une conscience des machines. Comme la nôtre, cette conscience émerge d'une réflexion entre deux plans de réalité mais, pour le reste, elle est complètement différente. Est-il besoin de rappeler que nous fabriquons les machines concrètes, que nous programmons les machines abstraites, que nous planifions la correspondance entre les deux ordres de complexité et que nous produisons, sélectionnons, et étiquetons les données d'entraînement des modèles de langue? Le golem computationnel n'est ni autoreproducteur et auto-poïétique comme les organismes, ni sensible comme les animaux et son comportement probabiliste ne peut être confondu avec la marge d'autonomie que confère la pensée discursive humaine. Même si de nombreux acteurs veulent créer une intelligence artificielle générale autonome et dotée de sa propre conscience, le principal effet de leurs efforts est d'augmenter l'intelligence collective humaine en lui tendant le miroir – aussi imparfait, partial et partiel soit-il – d'un nouvel intellect agent.

8. L'humilité philosophique

Cette méditation serait incomplète sans une évocation des *limites* de la conscience discursive. À la fin du siècle des lumières, dans sa *Critique de la raison pure*, Kant découvre que la réflexion de la raison (la pensée discursive), livrée à elle-même, ne peut mener à une connaissance scientifique : sans concepts, la sensibilité est aveugle mais, en l'absence de sensibilité, les concepts – avec les raisonnements qui les enchaînent – sont vides.

Quant à la conscience, informée par l'ère du soupçon (Sarraute, 1956), une introspection élémentaire lui démontre son caractère fini, et même étroit.

Il faut donc situer un inconscient derrière l'horizon de la réflexion. Comment penser un inconscient qui, dès qu'il est connu, n'est justement plus inconscient? L'inconscient ne peut être que soupçonné à partir de signes étranges, illogiques, qui doivent être interprétés pour en former une image hypothétique. Le psychanalyste et philosophe Cornélius Castoriadis a baptisé « magmatique » (Castoriadis, 1975) le régime ontologique qui se dérobe à la conscience discursive ordinaire. En effet, la raison claire et distincte ne peut connaître qu'une réalité « ensembliste-identitaire ». Comme son nom l'indique, cette dernière suppose le principe d'identité (A = A à l'instant *t*) et la constitution d'ensembles et de sous-ensembles composés d'éléments atomiques. Une connaissance de l'inconscient, aussi paradoxale que paraisse cette expression, suppose une traduction du magmatique dans l'ensembliste-identitaire, les *deux* modes d'êtres étant trahis au passage. Je discerne un double inconscient : actuel et virtuel. L'inconscient *actuel* enveloppe les déterminations non réfléchies de la pensée présente. Cette irréflexion trouve son origine dans l'opacité du corps ; dans des nœuds émotionnels inextricables et douloureux ; dans l'histoire traumatique inassumée du sujet et de ses ancêtres ; dans les structures et les interdits culturels qui nous déterminent à notre insu ; dans les archétypes anthropologiques, schémas de représentation et d'interaction qui nous font agir sans que nous le sachions ; ou encore

dans la simple ignorance de notre ignorance. Nous n'en entrevoyons que des bribes épanouissantes, des signes incomplets, déformés, presque incompréhensibles à cause des déplacements, condensations, inversions, traductions imagées et mélanges narratifs que leurs formes subissent en se projetant sur la conscience discursive ordinaire. J'ajoute que cet inconscient actuel se distribue en degrés : il peut être complètement inaccessible, profondément enfoui mais encore atteignable ou bien proche de la surface. Il existe aussi un inconscient *virtuel*, celui du possible impensable, de la création de formes inouïes et des inimaginables catastrophes à venir. Et celui-ci, plus encore que l'autre, nous reste inaccessible, sinon par d'obscures intuitions. De nouveau, quelle est la nature de l'inconscient? Par ses effets existentiels, l'irréfléchi appartient au domaine dionysiaque : sombre, parcouru d'intensités sauvages, il nourrit un enthousiasme ou un abrutissement proche de la folie. Nous voici maniaques le soir dans l'ivresse du vin et mélancoliques le matin des suites de l'intoxication. Mais par son contenu caché, derrière le masque d'un magma opaque, je le soupçonne d'être apollinien et d'obéir à un ordre, à une musique supérieure que nous aimerions bien entendre.

Un sombre nuage d'inconnaissance encercle la conscience. Puisque la philosophie s'exerce à la conceptualisation réflexive, le même ténébreux magma la cerne aussi. Tout ce qu'elle peut en dire, outre l'interprétation risquée des signes déformés qui franchissent l'interface entre le clair et l'obscur, c'est qu'elle sait qu'elle ne le connaît pas. Les philosophies particulières se contentent d'explorer quelques horizons, rarement tous. Quant à la philosophie en général, comme n'importe quelle activité réflexive, elle ne parvient pas à expliciter l'ensemble de ses déterminations. Il ne s'agit pas ici de « briser l'orgueil de la raison » ni d'infliger une quelconque « blessure narcissique » à une humanité qui n'a pas besoin d'une nouvelle offense à son amour-propre, mais de cultiver une salutaire humilité.

Références bibliographiques

Al-Ghazâlî, A.-H. (2013). *La délivrance de l'erreur* (H. Boutaleb, Trans.). Al Buraq Éditions. (Original work published 11th–12th century)

Alexander of Aphrodisias. (2012). *On the soul: Part I: Soul as form of the body, parts of the soul, nourishment, and perception* (V. Caston, Trans.). Bristol Classical Press.

Aristote. (1977). *De l'âme* (J. Tricot, Trans.). Vrin. (Original work published 3rd century BCE)

Aristote. (1981). *La métaphysique* (Tome 2) (J. Tricot, Trans.). Vrin.

Averroès. (1999). *L'intelligence et la pensée* (A. de Libera, Trans.). Garnier-Flammarion. (Original work published 12th century)

Bergson, H. (1959a). *Essai sur les données immédiates de la conscience*. Dans *Œuvres*. Presses Universitaires de France. (Original work published 1889)

Bergson, H. (1959b). *Matière et mémoire*. Dans *Œuvres*. Presses Universitaires de France. (Original work published 1896)

Bloch, E. (2008). *Avicenne et la gauche aristotélicienne* (C. Maillard, Trans.). Éditions Premières Pierres.

Castoriadis, C. (1975). *L'institution imaginaire de la société*. Seuil.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Corbin, H. (1964). *Histoire de la philosophie islamique*. Gallimard.

Corbin, H. (1978). *En Islam iranien: Aspects spirituels et philosophiques* (4 vols., 2nd ed.). Gallimard.

Corbin, H. (1981). *La philosophie iranienne islamique aux 17e et 18e siècles*. Buchet/Chastel.

Corbin, H. (1999). *Avicenne et le récit visionnaire*. Verdier.

Davidson, H. A. (1992). *Al Farabi, Avicenna, and Averroes, on intellect: Their cosmologies, theories of the active intellect, and theories of human intellect*. Oxford University Press.

De Libera, A. (2005). *Métaphysique et noétique: Albert le Grand*. Vrin.

Dennett, D. (1991). *Consciousness explained*. Little, Brown and Co.

Descartes, R. (1953). *Œuvres et lettres* (A. Bridoux, Trans.). Gallimard.

Deutsch, E., & Dalvi, R. (Eds.). (2004). *The essential Vedanta: A new source book of Advaita Vedanta*. World Wisdom.

Eckhart, M. (2022). *Intégrale des 180 sermons* (L. Jouvet, Trans.). Almora. (Original work published late 13th–early 14th century)

Fichte, J. G. (2016). *La doctrine de la science* (É. Jalley, Trans.). L'Harmattan. (Original work published 1794)

Freud, S. (1967). *L'interprétation des rêves* (I. Meyerson, Trans.). Presses Universitaires de France. (Original work published 1926)

Hobson, J. A., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences, 23*(6), 793–842. https://doi.org/10.1017/S0140525X00003434

Husserl, E. (2011). *Recherches logiques* (Vols. 1–2) (D. Franck & P. Ricoeur, Trans.).

Gallimard. (Original work published 1900–1901)

Jankélévitch, V. (1981). *Le je ne sais quoi et le presque rien* (3 vols.). Seuil.

Maïmonide, M. (2012). *Le guide des égarés* (S. Munk, Trans.). Verdier. (Original work published 12th century)

Plotin. (1931). *Ennéades, Tome V: Cinquième Ennéade* (É. Bréhier, Trans.). Les Belles Lettres. (Original work published 3rd century)

Plotin. (1938). *Ennéades, Tome VI, 2e partie: Ennéade VI, 6–9* (É. Bréhier, Trans.). Les Belles Lettres. (Original work published 3rd century)

Sarraute, N. (1956). *L'ère du soupçon*. Gallimard.

Spinoza, B. (2023). *Éthique* (C. Appuhn, Trans.). Flammarion.

Takpo Tashi Namgyal. (1986). *Mahamudra, the quintessence of mind and meditation* (L. P. Lhalungpa, Trans.). Shambhala. (Original work published 16th century)

Thomas d'Aquin. (1999). *Contre Averroès* (A. de Libera, Trans.). Garnier-Flammarion. (Original work published 13th century)

Whitehead, A. N. (1978). *Process and reality: An essay in cosmology*. The Free Press.

Zadra, A., & Stickgold, R. (2021). *When brains dream: Understanding the science and mystery of our dreaming minds*. W. W. Norton & Company.