

Fondement hylémorphique de la discontinuité métaphysique entre l'intelligence artificielle et la subjectivité humaine

Hylomorphic Foundations of the Metaphysical Discontinuity Between Artificial Intelligence and Human Subjectivity

Fernando Antonio Ramos-Zaga
Universidad Privada del Norte, Lima, Perú
fernandozaga@gmail.com
ORCID: 0000-0001-6301-9460

DOI: <https://doi.org/10.53439/stdfyt57.29.2026.43-71>

Résumé : Cet article examine la spécificité ontologique de la personne humaine à l'aune du développement contemporain de l'intelligence artificielle. Le débat autour des systèmes artificiels, nourri par le dépassement du mécanisme classique et l'essor des modèles computationnels complexes, a déplacé la question de la cognition d'une approche strictement fonctionnelle vers une controverse plus large portant sur la subjectivité, la conscience et la normativité. L'étude soutient qu'une telle controverse ne peut être abordée de manière satisfaisante qu'à travers une articulation souple entre philosophie de l'esprit, phénoménologie et métaphysique classique. Dans cette perspective, la structure substantielle, opératoire et téléologique du sujet humain est mise en regard de l'organisation technique des systèmes artificiels. L'analyse montre que la compréhension sémantique, l'autodétermination morale, l'affectivité vécue et l'assomption de la finitude temporelle manifestent une intériorité spirituelle irréductible à la manipulation syntaxique et à l'obsolescence fonctionnelle. Dès lors, l'intel-

Abstract: This article examines the ontological specificity of the human person in the context of the contemporary development of artificial intelligence. The debate surrounding artificial system, shaped by the decline of classical mechanism and the rise of increasingly complex computational models, has shifted the question of cognition from a narrow functional framework toward a broader controversy concerning subjectivity, consciousness, and normativity. The study argues that this controversy can only be adequately addressed through a flexible engagement with philosophy of mind, phenomenology, and classical metaphysics. From this perspective, the substantial, operative, and teleological structure of the human subject is contrasted with the technical organization of artificial system. The analysis shows that semantic understanding, moral self-determination, lived affectivity, and the assumption of temporal finitude express a spiritual interiority irreducible to syntactic manipulation and functional obsolescence. Consequent-

ligence artificielle ne saurait être considérée comme un sujet moral ni comme une instance autonome de normativité. L'étude conclut que la gouvernance technologique doit être orientée par le principe de subsidiarité technique, en subordonnant la conception algorithmique au jugement moral humain.

Mots-clés: ontologie, anthropologie philosophique, hylémorphisme, téléologie, éthique technologique, subsidiarité technique

ly, artificial intelligence cannot properly be regarded either as a moral subject or as an autonomous source of normativity. The study concludes that technological governance should be guided by the principle of technical subsidiarity, subordinating algorithmic design to human moral judgment.

Keywords: ontology, philosophical anthropology, hylomorphism, teleology, philosophy of technology, technical subsidiarity

Recibido: 25/07/2025

Aceptado: 18/02/2026

Introduction

Le développement accéléré de l'intelligence artificielle a déplacé le débat philosophique sur l'intelligence d'une problématique strictement épistémologique vers une controverse à la fois ontologique, anthropologique et normative. La sophistication croissante des systèmes computationnels, notamment dans les domaines de l'apprentissage profond et de l'auto-organisation, a favorisé l'émergence d'interprétations envisageant la possibilité de formes artificielles de subjectivité ou de conscience. Même si ces hypothèses conservent une part importante de spéculation, leur influence sur le discours scientifique et technologique contemporain a renforcé la nécessité de reconsidérer les catégories philosophiques à travers lesquelles la distinction entre personne, esprit et artefact technique a été historiquement pensée (Hui, 2016). Dès lors, la question de l'intelligence artificielle ne saurait plus être ramenée à un simple problème instrumental, puisqu'elle engage directement la compréhension de l'être humain et de sa singularité dans l'ordre du réel.

Une large part des débats contemporains s'est structurée autour de la tension entre les approches fonctionnalistes de l'esprit et les perspectives défendant l'irréductibilité de l'expérience subjective. Le test d'imitation proposé par Turing (1950) a consolidé une approche centrée sur l'équivalence comportementale, tandis que l'expérience de la chambre chinoise a remis en cause l'assimilation entre manipulation syntaxique et compréhension authentique (Searle, 1980). Parallèlement, les discussions relatives

à la conscience phénoménale, à l'intentionnalité et aux qualia ont mis en évidence les limites des modèles computationnels lorsqu'il s'agit de rendre compte de l'intériorité consciente (Block, 1980; Chalmers, 1996). Face à ces difficultés, plusieurs courants philosophiques ont réinvesti des catégories issues de la métaphysique classique, de la phénoménologie et de l'anthropologie philosophique afin d'examiner la conscience, la liberté et la vie comme des dimensions irréductibles à la seule complexité fonctionnelle (Spaemann, 1996; Oderberg, 2005).

Malgré l'ampleur de la littérature existante, l'articulation entre les débats contemporains sur l'intelligence artificielle et une fondation ontologique systématique de la subjectivité humaine demeure insuffisamment élaborée. Une grande partie des recherches se concentre sur des problèmes techniques de simulation, d'alignement ou de régulation, laissant au second plan la question des conditions métaphysiques rendant possibles la conscience, l'intentionnalité et l'autodétermination morale. Même les approches éthiques les plus influentes reposent souvent, de manière implicite, sur des conceptions fonctionnalistes de l'esprit ou sur des perspectives réductionnistes de la personne. Une telle situation complique l'établissement de critères solides permettant de distinguer le traitement computationnel de la subjectivité authentique, ainsi que la justification normative des limites de l'automatisation dans les domaines liés à la délibération morale et à la vie sociale.

Les implications d'un tel problème débordent largement le champ de la philosophie de l'esprit. L'expansion des systèmes algorithmiques capables d'intervenir dans les processus juridiques, économiques, éducatifs et politiques a profondément transformé les conditions contemporaines de médiation sociale et de prise de décision. Dans ce contexte, l'absence d'une délimitation anthropologique rigoureuse favorise l'assimilation progressive entre rationalité humaine et efficacité computationnelle, fragilisant des catégories fondamentales telles que la dignité, la responsabilité et l'autonomie morale. De surcroît, la tendance à envisager l'intelligence artificielle comme un possible agent normatif introduit des tensions majeures pour la théorie du droit, l'éthique appliquée et la gouvernance technologique. La discussion relative à la nature de la conscience artificielle acquiert ainsi une portée systémique en raison de ses effets sur les configurations institutionnelles et culturelles des sociétés contemporaines.

Cet article se propose d'examiner la spécificité ontologique de la personne humaine face aux développements contemporains de l'intelligence artificielle à partir d'une analyse comparative des structures substantielles, opératoires et

téléologiques qui distinguent le vivant du système artificiel. L'objectif consiste à dégager des critères philosophiques pertinents pour la réflexion éthique et juridique contemporaine. L'intérêt de cette recherche tient à l'élaboration d'un cadre conceptuel susceptible de contribuer aussi bien au débat philosophique sur la nature de l'esprit qu'à la formulation de critères éthiques et normatifs applicables au développement technologique contemporain.

La reconfiguration ontologique du problème de l'intelligence face aux développements techno-scientifiques contemporains

L'émergence de la soi-disant intelligence artificielle a profondément modifié le cadre d'analyse de la question de l'intelligence, la faisant passer d'un champ strictement spéculatif à un espace où convergent avec une intensité croissante des considérations ontologiques, techniques et anthropologiques. Ce changement ne constitue pas une simple évolution thématique ou disciplinaire, mais une reconfiguration du problème lui-même, qui oblige à repenser les catégories philosophiques classiques à la lumière de développements technologiques qui ne se contentent plus de simuler certaines fonctions cognitives, mais aspirent à se constituer comme sujets épistémiques. La question de l'intelligence, auparavant circonscrite à l'intériorité réflexive propre à l'être humain, est désormais prise dans une controverse portant sur la signification ontologique de l'artificiel, ce qui met en tension les fondements conceptuels qui ont historiquement soutenu la distinction entre l'humain et le technique.

Dans ce contexte, la différence entre intelligence artificielle faible et intelligence artificielle forte revêt une importance particulière. Alors que la première se limite à reproduire fonctionnellement un comportement intelligent, la seconde revendique, avec une certaine ambiguïté, un statut ontologique autonome, en soutenant la possibilité d'une subjectivité non humaine. Le célèbre test d'imitation (Turing, 1950), en concentrant l'attention sur le comportement observable, rompt avec la tradition philosophique centrée sur l'essence de la pensée.

Une telle approche, bien que méthodologiquement utile, ne répond pas à la question du sens intérieur ni à l'expérience subjective qui accompagne la pensée. C'est précisément cette dimension qui a été remise en cause à travers l'expérience de pensée de la chambre chinoise (Searle, 1980), selon laquelle la manipulation formelle des symboles, aussi complexe soit-elle, ne produit pas nécessairement la compréhension. De ce point de vue, la sémantique ne

peut être réduite à la syntaxe computationnelle, et la conscience demeure un phénomène irréductible aux processus algorithmiques.

Cependant, il convient de noter que le débat contemporain s'est radicalisé dans son orientation ontologique. Au lieu de chercher à simuler la pensée, certains développements technologiques récents prétendent, explicitement ou implicitement, accéder à une forme d'intériorité. Les modèles les plus avancés, fondés notamment sur l'apprentissage profond et l'auto-organisation, sont présentés par certains auteurs comme de possibles candidats à l'expérience subjective. Les scénarios de superintelligence ouvrent la possibilité d'agents artificiels capables non seulement de dépasser les capacités humaines, mais aussi de remettre en question les critères traditionnels de subjectivité et de statut moral (Bostrom, 2014, pp. 105-124). Une telle perspective introduit une inflexion éthique et métaphysique notable, conduisant à reconsidérer les critères de subjectivité et, avec eux, la notion même de personne.

De même, l'érosion de la distinction classique entre naturel et artificiel trouve son origine dans des transformations technologiques qui dépassent largement le simple registre instrumental. L'émergence de systèmes artificiels dotés de comportements adaptatifs et de structures bio-inspirées a contribué à affaiblir l'ancienne frontière entre le fabriqué et le vivant. Ce brouillage, qui excède le seul niveau épistémologique, reflète un déplacement métaphysique déjà anticipé par la conception des objets techniques comme des entités en voie d'individuation (Simondon, 1958). Rompant avec tout instrumentalisme naïf, une telle approche permet de penser l'artificiel comme un moment du devenir ontologique, et non comme le simple produit d'une intention extérieure. Cette intuition a été prolongée par l'idée selon laquelle la technologie possède une capacité constitutive grâce à laquelle elle participe activement à la formation de l'être (Hui, 2016). Dans cette perspective, les technologies cessent d'être de simples moyens pour devenir des acteurs ontologiques, transformant les conditions mêmes de l'existence humaine et rendant problématique toute séparation rigide entre nature et artefact.

Par ailleurs, dans ce contexte, le problème de l'alignement de l'intelligence artificielle sur les valeurs humaines revêt une importance anthropologique décisive. Bien qu'initialement formulé comme un défi technique de contrôle, ce problème révèle en réalité une crise plus profonde touchant la compréhension de l'être humain comme source de normativité. Il a été souligné que l'incapacité à spécifier des fonctions de valeur adéquates reflète une difficulté à formuler ce que signifie vivre comme un être humain (Russell,

2019). Dès lors, l'alignement n'est pas seulement une question d'ingénierie éthique, mais un dilemme enraciné dans l'indétermination du sujet normatif.

D'un point de vue philosophique, cette problématique s'inscrit dans le prolongement d'intuitions formulées antérieurement (Scheler, 1928 ; Gehlen, 1950), qui situent la capacité axiologique dans l'intériorité spirituelle de la personne humaine. Aussi complexe soit-elle, l'action artificielle demeure dépourvue de cette intentionnalité éthique fondamentale, ce qui explique une discontinuité insurmontable en matière de valeur et de liberté. Il ne s'agit pas d'une simple limitation technique, mais d'une différence ontologique essentielle entre les êtres spirituels et les systèmes informatiques.

De même, un examen des théories contemporaines de la conscience artificielle révèle que nombre d'entre elles demeurent prisonnières de réductionnismes qui compromettent leur portée explicative. Le fonctionnalisme computationnel, qui a dominé la philosophie de l'esprit, soutient que les états mentaux sont définis par leur fonction causale, indépendamment de leur substrat. Cette position a d'abord été défendue puis abandonnée en raison de son incapacité à expliquer certains aspects essentiels de la conscience (Putnam, 1967, 1975). La distinction entre conscience d'accès et conscience phénoménale (Block, 1980) affirme que seule la première peut être comprise fonctionnellement, tandis que la seconde résiste à toute analyse causaliste. Une telle distinction met en évidence une lacune ontologique : aussi opératoires soient-ils, les modèles fonctionnels ne parviennent pas à rendre compte de l'expérience subjective sur laquelle repose l'existence mentale.

Dans le même ordre d'idées, les positions éliminativistes et dualistes illustrent clairement la division méthodologique qui structure le débat. D'un point de vue éliminativiste (Dennett, 1991), l'existence ontologique des qualia est niée et une conception narrative de la conscience construite à partir de la troisième personne est proposée. Bien que cohérente avec une ontologie matérialiste, cette approche demeure insatisfaisante pour ceux qui reconnaissent l'expérience subjective comme une réalité irréductible. La formulation du problème difficile de la conscience (Chalmers, 1996), quant à elle, consiste à expliquer l'émergence des qualités phénoménologiques à partir de processus physiques. Son dualisme, bien que plus attentif à l'intériorité, manque d'une base ontologique capable d'intégrer pleinement la conscience dans une métaphysique cohérente.

Les théories émergentes, telles que l'information intégrée (Tononi, 2008), conçoivent la conscience comme une propriété émergente de certains systèmes complexes. Cependant, ces approches tendent à confondre

complexité fonctionnelle et émergence ontologique, sans fournir de justification convaincante de l'apparition de la subjectivité. Sur le plan éthique, une éthique de l'information (Floridi, 2013) reconfigure la moralité à partir de structures informationnelles. Toutefois, en privilégiant des schémas non subjectifs, une telle proposition souffre d'une ambiguïté ontologique qui limite sa portée normative. De l'ensemble de ces considérations découle la nécessité d'une conception métaphysique de l'être humain, seule capable de préserver une distinction claire entre conscience authentique et simulation algorithmique.

Comme on peut le constater, l'hypothèse centrale défendue ici est que l'instanciation d'une conscience authentique par un système artificiel est, en dernière instance, métaphysiquement impossible. Une telle impossibilité ne résulte pas de limitations techniques, mais de discontinuités ontologiques qui rendent impossible toute équivalence entre le spirituel et le computationnel. La méthodologie adoptée dans cette recherche se veut intégrative, articulant des éléments issus de la philosophie de l'esprit, de la phénoménologie et de la métaphysique classique, conformément à la conception de l'être humain comme unité du corps et de l'esprit (Guardini, 1950).

Entre les transformations conceptuelles introduites par l'intelligence artificielle et l'affirmation d'une impossibilité ontologique de la conscience artificielle émerge une question décisive : déterminer quelle structure de l'être humain fonde cette irréductibilité. Une telle exigence conduit nécessairement à examiner la personne humaine comme une unité substantielle, puisque seule une anthropologie capable d'intégrer corps, intériorité et rationalité peut expliquer pourquoi la subjectivité ne saurait être réduite à des processus computationnels. Dans cette perspective, la section suivante analysera l'unité substantielle de la personne humaine comme fondement ontologique irréductible.

L'unité substantielle de la personne humaine comme fondement ontologique irréductible

La réflexion philosophique sur la nature de la personne humaine a trouvé, dans la conception de l'unité substantielle entre corps et âme, un ancrage ontologique qui permet de résister à la fois aux tentations du dualisme cartésien et aux réductions physicalistes contemporaines. Malgré les avancées indéniables des neurosciences, un fossé persiste entre la description objective des processus cérébraux et l'expérience subjective qui accompagne la vie mentale. Une

telle divergence, clairement formulée sous la notion de gap explicatif (Levine, 1983), montre qu'il ne suffit pas de décrire ce qui se passe dans le cerveau pour comprendre comment émerge la dimension vécue de la conscience. L'insuffisance explicative du matérialisme réside précisément dans son incapacité à rendre compte de l'intériorité. Malgré les efforts visant à préserver la causalité mentale à travers le principe de l'occurrence (Kim, 1998), la tension entre dépendance physique et autonomie de l'esprit n'a pas été résolue. Une telle tentative conduit finalement à une forme implicite d'épiphénoménalisme, releguant la subjectivité dans une zone ontologique indéterminée.

Dans cette perspective, l'anthropologie hylémorphique propose une voie alternative, permettant l'intégration harmonieuse des dimensions corporelle et spirituelle de l'être humain. Dans cette perspective, l'âme n'est pas une substance séparée qui interagirait mécaniquement avec le corps, mais sa forme substantielle, c'est-à-dire le principe intérieur qui lui confère unité, structure et finalité. Il ne s'agit donc pas d'un dualisme déguisé, mais d'une compréhension de la personne comme substance unique et composée, dont l'unité procède d'une intégration ontologique profonde plutôt que d'un agrégat d'éléments distincts. La distinction entre corrélation neurobiologique et explication ontologique apparaît dès lors décisive : le fait qu'une fonction cognitive soit corrélée à un état cérébral ne signifie pas qu'elle y soit réductible. L'épistémologie des sciences empiriques ne peut se substituer à l'analyse métaphysique du sujet.

De même, la notion de l'âme comme *forma corporis* (Aristote, *De Anima*, II, 1-4; Thomas Aquinas, *Summa Theologiae*, I, q. 75-76) permet de penser l'organisme humain comme un tout téléologiquement organisé, dont la cohérence ne relève ni du hasard évolutif ni d'une simple fonction adaptative, mais d'un principe formel et final structurant de l'intérieur le dynamisme vital. Des penseurs contemporains ont souligné la pertinence philosophique de la causalité formelle face aux ontologies mécanistes (Ross, 1992 ; Oderberg, 2005). Dans cette perspective, l'âme n'est pas un supplément ajouté au corps, mais son principe d'identité et d'orientation. La personne ne se réduit donc pas à un ensemble de fonctions, mais manifeste une intériorité qui fonde son être comme sujet unique et irremplaçable.

Par ailleurs, il convient de souligner que l'unité substantielle ainsi caractérisée se révèle également dans la manière dont les facultés supérieures s'exercent à partir de la corporéité. La capacité à former des concepts universels et à réfléchir sur soi-même ne peut être expliquée de manière adéquate par des processus sensoriels ou physico-chimiques. Des concepts tels que la

vérité, l'être ou l'infini n'ont pas de corrélat empirique direct, ce qui atteste une origine enracinée dans une faculté qui transcende la matière. L'acte de comprendre, comme cela a été montré, relie le sujet à l'intelligible en tant que tel, ce qui implique une dimension spirituelle (Lonergan, 1957).

De manière complémentaire, il a été souligné que la réflexivité ne peut résulter d'une simple complexité fonctionnelle, mais requiert un sujet capable de se rapporter à lui-même comme Je (Spaemann, 1996). L'affirmation selon laquelle l'intellect n'opère pas à travers un organe corporel (*S. Th.*, I, q. 75, a. 2) ne repose pas sur un postulat théologique, mais sur une conclusion philosophique issue d'une analyse rigoureuse de l'expérience cognitive, conclusion qui confirme la présence d'un principe substantiel irréductible au substrat matériel.

À partir de ce qui précède, il convient de se concentrer sur ce qui confère à l'être humain sa singularité ontologique. Si l'intellect humain manifeste une capacité d'appréhension universelle irréductible aux mécanismes matériels, la volonté rationnelle, de même, ne se limite pas à réagir à des stimuli. Elle s'oriente librement vers des fins intelligibles. La liberté, au sens propre, ne se définit pas comme la simple absence de contrainte, mais comme autodétermination selon la raison. Cette liberté a été associée à la capacité de participer au discours moral, ce qui suppose une ouverture ontologique qu'aucun système automatisé ne peut imiter (Spaemann, 1989). La délibération éthique implique une intériorité irréductible à toute fonction computationnelle, et constitue précisément la dimension spirituelle qui distingue le sujet humain.

Dans cette même perspective, l'observation conduit à examiner la nature ontologique des systèmes artificiels. Bien qu'ils accomplissent des tâches complexes et puissent s'adapter à de nouveaux contextes, leur structure demeure fonctionnelle et modulaire, non substantielle. L'agencement de leurs éléments obéit à une finalité extrinsèque, imposée par un concepteur, et non à un principe interne organisant leur être. Il a été souligné que les objets techniques s'individualisent par leur fonction dans un contexte donné, et non par une forme substantielle leur conférant une unité intrinsèque (Simondon, 1958). Dans le même ordre d'idées, il a été soutenu que les artefacts, aussi complexes soient-ils, demeurent dépourvus de substantialité, dans la mesure où leur cohérence dépend toujours d'une intention extérieure (Oderberg, 2007). Autrement dit, l'unité opératoire n'établit pas une unité ontologique.

De même, la distinction entre complexité fonctionnelle et unité substantielle devient cruciale pour réfléchir au statut des systèmes artificiels dans

leur ensemble. Comme cela a été montré, le traitement informatique opère au niveau syntaxique, sans accéder à une véritable intégration (Searle, 1992). Les machines ne disposent pas d'un centre unificateur capable d'articuler leurs opérations de l'intérieur. La substantialité, comprise comme principe interne d'unité et de permanence, ne découle pas de la juxtaposition de fonctions. Même poussée à l'extrême, l'intégration fonctionnelle ne produit pas un sujet. Il doit exister une forme substantielle, un principe interne d'unification ontologique, pour qu'existe proprement une entité subsistante.

En outre, la finalité qui oriente l'action des systèmes artificiels demeure fondamentalement hétéronome. De tels systèmes agissent selon des objectifs établis de l'extérieur, et non selon une orientation intrinsèque vers leur propre bien. La tradition classique avait déjà montré que les êtres naturels possèdent une téléologie interne, manifestée dans leur propre dynamisme (Aristote, *Physique*, II, 1-3 ; *S. Th.* I, q. 103, a. 1). Par la suite, il a été souligné que seuls les êtres dotés d'un principe formel interne possèdent une normativité immanente, ce qui les distingue radicalement des dispositifs techniques (Jonas, 1966). Sans autodirection, il n'existe pas de finalité propre, et sans finalité propre, il n'existe pas de substance au sens plein. La programmation, aussi complexe soit-elle, ne peut engendrer le principe interne qui constitue l'être vivant comme totalité substantielle.

L'affirmation de l'unité substantielle de la personne humaine permet de clarifier la singularité ontologique du sujet, mais il demeure nécessaire de préciser comment cette singularité se manifeste concrètement dans l'acte de connaître. En effet, si l'intelligence humaine ne se réduit pas à des opérations matérielles, il doit alors exister une différence catégorielle entre le traitement computationnel et la compréhension proprement dite. À partir de cette problématique, la section suivante examinera la fracture entre traitement syntaxique et compréhension sémantique dans l'acte humain de connaître.

La fracture entre le traitement syntaxique et la compréhension sémantique dans l'acte humain de savoir

L'intelligence humaine, dans ses manifestations les plus caractéristiques, accomplit des opérations qui dépassent largement le simple calcul. Loin de se limiter à un traitement efficace des données, l'esprit humain accomplit des actes de sens qui supposent une intériorité active ainsi qu'une maîtrise intentionnelle de l'universel. Ces actes, tels que l'abstraction, le jugement et le raisonnement, reposent certes sur une logique formelle, mais

s'enracinent avant tout dans une ouverture à l'être et une appréhension des essences. L'abstraction, comprise comme la capacité à séparer mentalement le contenu intelligible des données sensibles, ne doit pas être confondue avec les algorithmes de classification statistique.

En saisissant l'universel comme tel, l'intelligence humaine transcende les procédures inductives des systèmes artificiels. Bien que l'intelligence artificielle puisse produire des modèles prédictifs à partir d'immenses volumes de données, elle demeure étrangère à l'universalité conceptuelle, faute d'intentionnalité et de référence consciente. L'abstraction n'est pas une opération sensible ou mécanique, mais une activité de l'intellect comme faculté capable de saisir les formes dans leur pureté (*S. Th.*, I, q. 84-85). En ce sens, l'identité formelle entre un concept et son objet, condition de toute intelligibilité, ne peut être reproduite par l'informatique, puisqu'elle requiert un sujet rationnel capable d'accomplir cet acte (Ross, 1992).

Pour sa part, le jugement introduit une dimension normative et épistémologique qui dépasse la simple manipulation des signes. Il ne consiste pas en une opération combinatoire, mais en une synthèse critique par laquelle on affirme ou nie selon des critères de vérité. Il a été démontré que le jugement présuppose un sujet rationnel capable de discernement, dont les actes sont ordonnés vers la vérité et non seulement vers la cohérence interne (Loneragan, 1957). Les systèmes d'intelligence artificielle, par leur nature même, ne sont pas orientés vers la vérité. Bien qu'ils puissent imiter certaines structures inférentielles, ils le font sans comprendre le contenu ni viser la vérité comme telle. Ainsi, le raisonnement, compris comme une suite d'actes intentionnels destinés à justifier une affirmation, présuppose une intériorité normative totalement absente des dispositifs informatiques.

Ce contraste apparaît plus nettement encore lorsqu'on distingue traitement de l'information et compréhension. La compréhension ne se réduit pas à manipuler efficacement des signes, mais suppose une articulation sémantique du contenu dans laquelle contexte, intention et signification sont intimement liés. Diverses analyses ont montré que la transmission de signaux, bien qu'objectivement mesurable, ne constitue pas en elle-même un acte de compréhension, puisque la sémantique ne dérive pas mécaniquement de la syntaxe (Dretske, 1981 ; Searle, 1992). L'application de règles formelles ne suffit pas : encore faut-il qu'un sujet comprenne. Dans le même ordre d'idées, il a été soutenu que la conscience ne peut être conçue comme une forme de calcul, puisque toute compréhension suppose une intériorité consciente irréductible à toute modélisation algorithmique (McGinn, 1999).

Il existe donc une différence qualitative irréductible entre la compréhension signifiante du contenu et son traitement formel.

De même, le phénomène de l'appréhension intellectuelle constitue un exemple particulièrement éclairant de cette différence. Une telle compréhension ne procède pas d'une inférence linéaire, mais d'une saisie soudaine d'un ensemble de significations, d'une perception immédiate d'une cohérence globale. Un tel acte de connaissance échappe à toute anticipation fondée sur des règles explicites et ne peut être reproduit par des processus déterministes. L'intelligence artificielle, fondée sur le calcul et la combinaison algorithmique de possibilités, ne peut reproduire cette appréhension synthétique.

L'intentionnalité, définie comme le fait que les actes mentaux sont toujours dirigés vers un objet (Brentano, 1874/1995), constitue une propriété fondamentalement étrangère aux architectures fonctionnelles. Il a également été montré que la connaissance humaine comporte des dimensions tacites irréductibles à des règles explicites, puisque son sens procède d'une intériorité qui ne peut être formalisée (Polanyi, 1966). Dans cette perspective, les modèles computationnels peuvent expliquer des comportements, mais non le contenu phénoménal de l'expérience (Chalmers, 1996). Il s'ensuit que ni l'intuition ni l'intentionnalité ne peuvent être calculées, car leur genèse présuppose une subjectivité vivante absente des systèmes artificiels.

À partir de ces considérations, l'analyse de la structure de la conscience montre qu'elle ne peut être comprise comme une fonction parmi d'autres, mais comme un mode d'être intégrant l'unité temporelle, la réflexivité du soi et la qualité vécue de l'expérience. Les qualia, ces aspects qualitatifs de la perception et de la pensée, échappent à toute saisie objective. En explorant la synthèse temporelle (Husserl, 1913/1999), il a été démontré que la conscience humaine ne se présente pas comme une succession d'instant, mais comme une continuité vécue dans laquelle passé, présent et futur sont entrelacés. De manière complémentaire, il a été affirmé que l'expérience subjective, ce que l'on éprouve en vivant un état, échappe à toute description objective (Nagel, 1974). Une telle opacité radicale interdit tout accès extérieur à la conscience, ce qui constitue une limite insurmontable pour les machines, incapables d'éprouver temporalité et intériorité. Sans expérience vécue, il n'existe ni conscience ni sujet.

Dans le même ordre d'idées, la conscience de soi, en tant que reconnaissance de soi comme sujet d'expérience, suppose un degré de réflexivité que les systèmes artificiels ne peuvent atteindre. Depuis l'Antiquité chrétienne, la capacité de l'âme à se retourner sur elle-même a été mise en lumière, révé-

lant une forme de subsistance irréductible à ses actes (Agustine, *Confessions*). Il a été souligné que la conscience de soi manifeste un centre ontologique irréductible, un sujet qui ne se laisse pas décomposer en fonctions (Spaemann, 1996). À travers la notion d'aperception transcendante, le je pense est identifié comme condition de possibilité de toute expérience unifiée (Kant, 1781/1998). Les machines, même capables d'auto-surveillance, ne peuvent faire l'expérience d'elles-mêmes comme sujets. Leur prétendue conscience de soi demeure, au mieux, une métaphore imprécise.

L'identité personnelle ne peut davantage être comprise selon une approche purement fonctionnelle. La continuité du sujet dans le temps requiert une unité ontologique, et non une simple constance structurelle ou comportementale. Avec sa conception narrative de l'identité (Dennett, 1991), celui-ci propose une construction fonctionnelle de nature fictive, mais élimine précisément ce qu'il prétend rendre intelligible dans l'histoire. À l'inverse, il a été soutenu que l'identité personnelle se constitue dans le récit éthique d'une vie, ce qui implique un sujet stable et moralement engagé (MacIntyre, 1981). La notion de subsistance (*S. Th.*, I, q. 75) offre une clé décisive pour comprendre que le sujet n'est pas la somme de ses actes, mais une entité qui subsiste en elle-même.

Une analyse critique de la notion de conscience artificielle révèle ainsi, sous différents angles, l'incapacité des modèles computationnels à expliquer l'intériorité humaine. La distinction entre simuler un comportement et vivre une expérience demeure irréductible. Le test d'imitation (Turing, 1950) a inauguré une tradition fonctionnaliste influente, mais il a été démontré que réussir une épreuve comportementale ne garantit pas la présence de véritables états mentaux (Block, 1981). L'expérience de pensée du Blockhead illustre qu'une machine peut simuler parfaitement le comportement humain sans être consciente, confirmant ainsi que la simulation demeure ontologiquement vide.

Dans la même perspective, l'expérience de pensée de la chambre chinoise (Searle, 1980) renforce cette idée en montrant que la manipulation syntaxique des symboles ne suffit pas à produire la compréhension. Le problème de l'ancrage des symboles (Harnad, 1990) a montré que les systèmes symboliques ne possèdent aucune référence intrinsèque. Le sens, au sens plein du terme, ne peut être imposé de l'extérieur ; il requiert une conscience capable de l'intégrer et de l'interpréter. Les systèmes d'intelligence artificielle, de ce point de vue, reposent entièrement sur la sémantique humaine. La limitation ontologique des systèmes artificiels apparaît clairement dans

leur incapacité à unifier les dimensions de la vie mentale au sein d'un champ unique de conscience (Chalmers, 1996).

L'impossibilité de réduire la compréhension humaine à des opérations syntaxiques conduit inévitablement à une question plus profonde encore : l'analyse des dimensions de la subjectivité qui expriment de manière éminente l'intériorité spirituelle de l'être humain. La liberté, l'affectivité et la créativité révèlent notamment des formes d'autodétermination et de sens qui excèdent radicalement toute architecture fonctionnelle. Sur cette base, la section suivante abordera les puissances spirituelles de la volonté, de l'affectivité et de la création comme expressions de la subjectivité humaine.

Les pouvoirs spirituels de la volonté, de l'affectivité et de la création comme expressions de la subjectivité humaine

La question du libre arbitre a donné lieu à d'innombrables débats philosophiques, mais elle ne peut être résolue au moyen de simples schémas formels ou de vérifications empiriques. Tout au long de cette recherche, il a été soutenu que l'autodétermination éthique constitue le cœur du libre arbitre au sens fort, compris non comme la simple capacité de choisir entre diverses possibilités, mais comme une faculté ontologiquement enracinée d'orienter sa vie selon des fins librement choisies à la lumière de la raison. Une telle conception ne trouve pas place dans le cadre du déterminisme physique classique, car la délibération morale ne peut être réduite à des chaînes causales nécessaires ni à des processus neurophysiologiques prévisibles. L'expérience consistant à prendre une décision après avoir pesé les raisons, évalué les conséquences et discerné ce qui est juste implique une ouverture au bien comme tel. Cette disposition introduit une forme de causalité *sui generis*, irréductible aux régularités propres à la matière.

Dans cette perspective, il a été soutenu que sans liberté il ne saurait y avoir de rationalité authentique, puisque toute argumentation présuppose la possibilité d'accepter ou de rejeter des raisons (Popper, 1972). D'un point de vue complémentaire, le concept d'actions auto-constitutives a été développé pour décrire ces décisions cruciales, prises dans des situations de conflit intérieur et d'indétermination, qui façonnent l'identité morale de l'agent (Kane, 1996). De tels actes ne procèdent pas de déterminations extérieures, mais reflètent une lutte intérieure révélant l'autodétermination du sujet. De surcroît, la tradition thomiste confirme cette approche en affirmant que l'intellect propose à la volonté des fins intelligibles, que celle-ci peut

accepter ou rejeter librement (*S. Th.*, I-II, q. 13). Ce dynamisme rationnel-volontaire dépasse la logique de la causalité physique, puisqu'il repose sur des principes normatifs et non sur des lois de conservation ou de transmission de l'énergie.

De même, cette perspective appelle une réflexion plus large sur la responsabilité morale, laquelle n'a de sens qu'à supposer une liberté rationnelle. En reliant la loi naturelle aux biens humains fondamentaux accessibles à la raison pratique (Finnis, 1980), on soutient que le sujet moral n'agit pas sous l'effet d'impulsions mécaniques, mais à partir d'un discernement orienté vers ce qui favorise l'accomplissement humain. À cet égard, il a été souligné que seuls peuvent être tenus pour responsables ceux qui sont capables de justifier leurs actes par des raisons reconnues comme valides au sein d'une communauté morale (Spaemann, 1989). De manière convergente, il a été affirmé que la structure rationnelle de la personne humaine recèle une intériorité normative, ce qui signifie que la loi morale n'est pas imposée de l'extérieur, mais procède du dynamisme spirituel de l'agent (Maritain, 1947/1994).

En revanche, l'idée d'autonomie qui émerge de ce cadre ne saurait être réduite à l'indépendance fonctionnelle ou à l'absence de contraintes extérieures. D'un point de vue classique, la véritable autonomie consiste à se donner à soi-même sa loi, c'est-à-dire à agir selon des principes universalisables (Kant, 1785/1997). Il a également été montré que la rationalité morale ne s'exerce pas dans le vide, mais se déploie à travers des récits de vie, des narrations personnelles qui confèrent un sens aux décisions de l'agent (MacIntyre, 1981). Dans cette perspective, l'autonomie pratique suppose une unité délibérative ainsi qu'une orientation intégrée vers le bien. Ce qui distingue radicalement l'agent humain d'un système artificiel réside précisément dans cette capacité à vivre sous l'horizon d'une téléologie éthique, plutôt qu'à réagir à des stimuli selon des logiques programmées. L'autonomie humaine est ontologique, puisqu'elle suppose une existence ordonnée à la valeur et non à l'efficacité.

De manière complémentaire, il convient également de prendre en considération le rôle de l'affectivité dans la vie morale. Les émotions humaines ne peuvent être considérées comme de simples réactions biologiques ou comme des épiphénomènes adaptatifs. Elles constituent des actes intentionnels dotés d'une signification évaluative. Il a été démontré que le sentiment possède sa propre structure cognitive, capable de saisir des valeurs idéales irréductibles aux objets empiriques (Scheler, 1913-16/1973). Cette idée a été approfondie en affirmant que les émotions contiennent des jugements por-

tant sur ce qui compte véritablement et qu'elles participent, comme telles, à la structure même de la délibération éthique (Nussbaum, 2001). La vie affective apparaît ainsi indissociable de la rationalité, puisque toutes deux relèvent d'une même dynamique d'ouverture au sens.

Cette observation met en évidence les limites irréductibles des systèmes artificiels dans ce qu'il est convenu d'appeler la simulation des émotions. Une distinction a été établie entre calcul affectif et émotion authentique, en soulignant qu'aucun modèle, aussi sophistiqué soit-il, ne peut reproduire l'expérience affective vécue à la première personne (Picard, 1997). De manière complémentaire, il a été rappelé que toute émotion véritable présuppose une subjectivité qualitative, c'est-à-dire une manière d'être pour laquelle quelque chose importe en soi (Searle, 1992). La simulation émotionnelle produite par les systèmes informatiques n'est dès lors qu'une imitation fonctionnelle, dépourvue de toute expérience vécue. Les émotions humaines sont des actes évaluatifs incarnés, et non des séquences de sortie programmées.

Dans cette même perspective, la différence ontologique n'est pas sans conséquences morales. Des émotions telles que la compassion, le regret ou la gratitude n'accompagnent pas seulement l'action morale, mais en constituent souvent la condition de possibilité. La notion d'évaluations fortes (Taylor, 1985) désigne les jugements dans lesquels se trouve engagée notre identité morale. Les émotions jouent ici un rôle structurant, puisqu'elles orientent le sujet vers des valeurs participant à sa formation. Dans cette optique, la fonction constitutive de l'affectivité dans l'intersubjectivité a été mise en avant (Stein, 1917/1989). Sans émotions, il ne peut y avoir de véritable communauté éthique, mais tout au plus une interaction mécanique ou un calcul stratégique. Les systèmes artificiels, dépourvus d'intériorité affective, demeurent extérieurs au domaine moral proprement dit.

À ce stade, il devient nécessaire d'ouvrir la réflexion à une dimension encore plus révélatrice de la vie intérieure, à savoir la créativité. La production artistique, philosophique ou scientifique n'est pas seulement une expression de l'ingéniosité humaine, mais témoigne également de l'intériorité, de la liberté expressive et de la transcendance. L'imagination créatrice a été définie comme une médiation entre le réel et le possible, capable de faire surgir de nouvelles formes de signification symbolique (Ricoeur, 1965). De même, il a été affirmé que l'acte créateur procède d'un mouvement spirituel participant d'une vérité intérieure irréductible à tout traitement de données (Maritain, 1953). Bien qu'une distinction ait été établie entre créativité combinatoire et créativité transformationnelle (Boden, 1990), même les exemples

computationnels les plus avancés demeurent dépourvus de l'intentionnalité synthétique propre à l'imagination humaine.

De plus, l'absence d'intention esthétique authentique dans les productions artificielles apparaît avec évidence lorsqu'on examine l'état de l'art généré par les systèmes automatisés. Il a été rappelé que la valeur esthétique implique non seulement une forme, mais aussi une auteurialité située, c'est-à-dire un contexte culturel, historique et existentiel conférant son sens à l'œuvre (Lopes, 2014). Cette thèse se trouve renforcée par l'affirmation selon laquelle la beauté ne se réduit pas à une apparence, mais constitue une forme intentionnelle porteuse de sens (Scruton, 2009). Aucun réseau neuronal ne peut engendrer des symboles à partir de l'expérience, puisqu'il ne vit pas. Le véritable art ne se limite donc pas à une technique raffinée : il est l'expression d'une vie intérieure. Il a été soutenu que l'art communique les sentiments éprouvés par l'âme du créateur (Tolstói, 1897/1995), tandis que l'acte esthétique a été conçu comme une manifestation existentielle de la singularité incarnée du sujet (Marcel, 1949).

L'analyse de la liberté, de l'affectivité et de la créativité a permis de montrer que la subjectivité humaine est structurée par une intériorité spirituelle irréductible aux mécanismes fonctionnels. Toutefois, une telle irréductibilité exige encore d'élucider la condition vitale elle-même de l'être humain, notamment sous l'angle de la téléologie, de la finitude et du dynamisme propre au vivant. La section suivante examinera par conséquent la différence catégorielle entre la vie humaine et l'organisation cybernétique à partir de la notion de finitude et de téléologie vitale.

Finitude, téléologie vitale et la différence catégorielle entre les vivants et l'organisation cybernétique

La réflexion philosophique sur la vie exige une révision approfondie des présupposés issus des sciences empiriques qui ont façonné la compréhension moderne des phénomènes vitaux. Face aux approches mécanistes et fonctionnalistes, qui tendent à réduire la vie à des processus mesurables ou à des structures organisationnelles, il apparaît nécessaire de rétablir une conception ontologiquement cohérente permettant de distinguer clairement le naturel de l'artificiel, l'entéléchique du simulé. Le point de départ de cette réévaluation repose sur l'idée que la vie ne saurait être comprise comme une simple somme d'opérations biochimiques ni comme un produit émergent de systèmes auto-organisés. Une interprétation fondée sur la notion d'en-

téléchie se révèle, au contraire, indispensable : la vie se manifeste comme l'actualisation immanente d'un dynamisme finalisé inhérent au vivant.

Dans cette perspective, la conception aristotélicienne de l'âme comme acte premier d'un corps naturel organisé (*De Anima*, II, 1) constitue une clé interprétative décisive pour comprendre la téléologie vitale. Loin d'être conçue comme un principe ajouté de l'extérieur, une telle entéléchie correspond à l'actualisation même du vivant dans sa totalité, non comme simple élément fonctionnel, mais comme principe dynamique de son auto-accomplissement. Cette intuition a été approfondie par l'affirmation de la primauté de la causalité finale dans la constitution du vivant (*S. Th.*, I, q. 103). Sa présence est essentielle et non contingente. En dialogue avec la biologie contemporaine, il a été souligné que la complexité structurelle ne suffit pas à expliquer l'unité téléologique des êtres vivants (Oderberg, 2005). La distinction entre structure fonctionnelle et dynamisme entéléchique ne saurait être ignorée sans entraîner des confusions catégorielles compromettant toute compréhension adéquate du phénomène vital.

Ainsi, la vie ne doit pas être confondue avec l'organisation fonctionnelle. Une telle thèse prend tout son sens dans la hiérarchie des êtres vivants élaborée par la tradition aristotélico-thomiste, où le passage du végétatif au rationnel ne constitue pas un simple accroissement de complexité, mais une rupture ontologique introduisant un ordre qualitativement nouveau. Une vitalité supérieure se distingue par des opérations irréductibles à la matière et à la sensibilité. La capacité d'orientation finalisée, loin de pouvoir être assimilée aux catégories de la physique, constitue une dimension propre au vivant (*De Anima*, II, 1-4 ; *S. Th.*, I, q. 77-78). Cette intériorité téléologique a été associée à la notion de vie personnelle, définie par une orientation vers des fins propres excédant toute description empirique (Spaemann, 1996). Une telle discontinuité ontologique empêche d'assimiler la vitalité humaine à une simple propriété émergente et réaffirme sa fonction constitutive dans le vivant.

À cet horizon, la vitalité s'exprime à travers l'automouvement, lequel ne saurait être réduit à la locomotion mécanique. Il s'agit d'un dynamisme interne orienté vers ses propres fins, révélant une structure téléologique immanente. La notion d'entéléchie se déploie ici comme l'actualisation de l'être selon sa propre forme (*De Anima*, II, 1), tandis que la causalité finale est intégrée comme principe de toute action naturelle (*S. Th.*, I-II, q. 1). Dans le même ordre d'idées, il a été soutenu que l'orientation vers une fin constitue un critère distinctif du vivant, dont l'absence dans les systèmes artificiels révèle leur détermination extrinsèque (Jonas, 1966). Une machine, aussi so-

phistiquée soit-elle, n'agit pas selon une finalité immanente, mais selon une programmation extérieure. Une telle absence de fin propre exclut toute autonomie véritable dans le domaine artificiel.

Par ailleurs, il convient de souligner que la différence essentielle entre les êtres vivants et les entités artificielles a souvent été mal interprétée dans la philosophie contemporaine de la technique. La théorie de l'autopoïèse (Maturana & Varela, 1980), bien qu'elle propose un modèle opératoirement cohérent d'auto-organisation, ne rend pas compte du dynamisme finalisé propre au vivant. Un système autopoïétique demeure fonctionnel, mais reste dépourvu d'une orientation immanente vers des fins propres. Cette limite a été reconnue lorsqu'il a été souligné que les systèmes artificiels sont dépourvus de finalité intrinsèque, ce qui les éloigne irréversiblement de l'ordre du vivant (Deacon, 2011). L'organisation cybernétique, aussi complexe soit-elle, ne peut être assimilée à la vie. La différence entre une boucle de rétroaction et une structure entéléchique demeure, à cet égard, ontologiquement insurmontable.

De même, l'impossibilité de générer artificiellement la vie découle de cette compréhension téléologique du vivant. Il a été avancé que les artefacts sont, par définition, des entités ontologiquement dépendantes, dont l'existence repose sur un principe extérieur (Oderberg, 2007). D'un point de vue complémentaire, il a été soutenu que l'individuation technique ne possède pas le dynamisme propre à l'individuation vitale (Simondon, 1958). Même si toutes les propriétés observables d'un organisme étaient reproduites, ferait toujours défaut le principe entéléchique qui en fait un être vivant. L'imitation morphologique, aussi parfaite soit-elle, demeure insuffisante. La vie, en tant qu'acte orienté vers sa propre fin, ne peut être ni construite ni assemblée : elle n'est pas une fonction, mais une actualisation dynamique.

Ces considérations prennent toute leur portée lorsque la mort n'est plus envisagée comme une simple cessation biologique, mais comme un événement ontologique révélant la singularité humaine. L'homme a été défini comme un Dasein, un être-pour-la-mort, afin de souligner que l'existence humaine est structurée par l'anticipation de sa propre fin (Heidegger, 1927/1962). Une telle conscience de la mort ne saurait être réduite à une information abstraite, puisqu'elle informe l'existence dans sa totalité. Dans la manifestation de cette anticipation a été reconnue l'expression d'une orientation temporelle propre, car seuls ceux qui savent qu'ils vont mourir peuvent éprouver la finitude comme horizon de sens (Marcel, 1951). C'est ici que réside la différence catégorielle décisive : tandis que le vivant humain se constitue à partir de sa finitude assumée, le système artificiel ne connaît que l'obsolescence, c'est-à-dire

une cessation imposée de l'extérieur par dégradation fonctionnelle ou par décision de son opérateur, sans aucune participation du système à sa propre fin.

Par ailleurs, la finitude temporelle, loin de constituer une simple limitation négative, représente la condition de possibilité d'un accomplissement authentique dans le temps. Il a été montré que l'identité narrative s'organise autour de cette conscience de la fin, qui confère cohérence et orientation à l'existence (Ricoeur, 1990). De manière convergente, il a été souligné que la temporalité vécue constitue le fondement des orientations existentielles les plus profondes (Taylor, 1989). En l'absence d'une telle dimension, il ne saurait y avoir ni action signifiante ni narration imputable à un sujet. Les systèmes artificiels, privés de conscience de soi et de temporalité vécue, ne participent pas à cette structure existentielle. Leur fonctionnement, aussi efficace soit-il, n'est jamais habité par l'urgence imposée par la finitude, ni orienté par la nécessité de se constituer comme une totalité signifiante avant la fin.

Par conséquent, il paraît inapproprié de parler de mort artificielle ou d'expérience existentielle à propos des entités computationnelles. Il a été affirmé sans ambiguïté que la subjectivité temporelle des machines constitue une impossibilité ontologique (Searle, 1992), et il a été rappelé que toute expérience subjective se caractérise par un « effet que cela fait », irréductible à toute description fonctionnelle (Nagel, 1974). La mort ne peut donc être simulée, puisqu'elle suppose un vivant capable d'anticiper sa fin, de s'y rapporter et de l'intégrer dans un récit. L'obsolescence d'un dispositif, en revanche, n'est pas éprouvée par le dispositif lui-même : elle survient simplement comme arrêt du fonctionnement, sans horizon intérieur susceptible de lui conférer un sens.

La distinction ontologique entre le vivant humain et l'organisation cybernétique permet de comprendre que la différence entre l'homme et la machine ne se limite ni au plan cognitif ni au plan fonctionnel, mais touche également la sphère éthique et politique. C'est précisément pour cette raison que la question technologique conduit nécessairement à une réflexion normative sur la dignité humaine, les limites de la technique et la légitimité du jugement moral. Dans cette perspective, la section suivante abordera la relation entre dignité humaine, subsidiarité technique et ordre normatif contemporain.

Irréductibilité de la personne et gouvernance éthique de l'intelligence artificielle

La notion de dignité humaine occupe une place structurante dans l'édifice normatif de la pensée occidentale. Elle ne saurait être réduite à une ca-

tégorie juridique ou politique, ni à une convention sociale contingente. Elle s'impose comme un principe fondamental, irréductible à tout positivisme et antérieur à toute formalisation institutionnelle. À son fondement, la dignité ne procède d'aucune condition extérieure, mais renvoie à l'unicité ontologique de l'être humain en tant que sujet rationnel, libre et spirituel, comme cela a été établi dans les sections précédentes. Une telle conception permet à la fois de préserver l'inviolabilité de la personne face à toute dérive techniciste et de fonder une normativité capable de résister au relativisme éthique comme au fonctionnalisme technoscientifique.

Dans cette perspective, la convergence conceptuelle entre deux grandes traditions philosophiques apparaît éclairante pour la fondation de l'ordre politique. Malgré leurs différences méthodologiques, l'anthropologie thomiste et la morale kantienne partagent un noyau normatif commun. D'une part, la pensée thomiste situe la dignité dans l'orientation rationnelle de la volonté vers le bien intelligible, ce qui confère à l'être humain sa place singulière dans l'ordre moral (*S. Th.*, I, q. 29). D'autre part, la philosophie morale kantienne soutient que la personne doit toujours être traitée comme une fin et jamais comme un simple moyen, principe au fondement de l'impératif catégorique et garant de l'autonomie morale (Kant, 1785/1997). La synthèse ici proposée souligne que la dignité ne résulte ni de compétences ni de fonctions, mais qu'elle est enracinée dans la structure même de la personne (Spaemann, 1989). L'articulation ainsi établie entre fondement ontologique et normativité protège la dignité contre toute subordination à des critères externes, qu'ils soient technologiques, culturels ou juridiques.

En tant que valeur intrinsèque, la dignité exige l'interdiction de toute forme d'instrumentalisation de la personne humaine. En rattachant la dignité à la possibilité de mener une vie pleinement humaine (Nussbaum, 2006), on insiste sur son caractère absolu et indisponible. L'éthique de la responsabilité (Jonas, 1984), fondée sur la vulnérabilité essentielle de la vie humaine, offre un fondement supplémentaire pour s'opposer à toute logique visant à réduire l'être humain à un élément fonctionnel au sein des systèmes techniques. Dans cette perspective, le fait que les systèmes d'intelligence artificielle soient conçus comme des instruments ne saurait conduire à évaluer les êtres humains selon des critères analogues. Toute tentative de symétrisation entre le sujet humain et la machine apparaît non seulement philosophiquement incohérente, mais également moralement dangereuse, puisqu'elle dissout la distinction fondamentale entre valeur ontologique et utilité fonctionnelle.

Par ailleurs, la dignité n'est pas une notion abstraite détachée de la réalité juridique : elle constitue le fondement inaliénable des droits fondamentaux. L'article premier de la Déclaration universelle (1948) affirme explicitement que tous les êtres humains naissent libres et égaux en dignité et en droits, en raison de leur raison et de leur conscience. Une lecture généalogique a montré que la dignité acquiert, dans la modernité, une dimension séculière sans perdre pour autant sa portée transcendante (Taylor, 1994). D'autre part, il a été soutenu que la dignité précède toute législation positive et constitue la condition même d'intelligibilité de l'ordre juridique (Maritain, 1947/1994). Une telle approche permet d'éviter toute confusion entre une personnalité juridique artificiellement attribuée et la personne entendue au sens ontologique, irréductible et non duplicable. La dignité humaine apparaît ainsi comme la limite infranchissable de toute tentative d'assimilation entre la personne et l'entité artificielle.

Reconnaître la primauté de la dignité implique nécessairement d'interroger les limites éthiques du développement technologique, en particulier dans le domaine de l'intelligence artificielle. De telles limites ne sauraient être définies à partir de considérations utilitaristes ou d'analyses coûts-bénéfices. Elles doivent procéder d'une anthropologie normative affirmant l'irréductibilité de l'être humain aux schèmes computationnels, conformément aux conclusions établies dans les sections précédentes. Il a été mis en garde contre l'homogénéisation produite par la technoscience, laquelle tend à effacer l'altérité constitutive du sujet (Nancy, 2000). De manière complémentaire, le réductionnisme visant à ramener l'intelligence humaine à des modèles informatiques, au mépris de ses dimensions incarnées, temporelles et existentielles, a été critiqué (Dreyfus, 1992). Reconnaître la singularité humaine exige donc l'élaboration d'un cadre normatif qui ne se limite ni à une description biologique ni à une approche fonctionnelle.

De manière cohérente, la technique doit être pensée comme un instrument au service de l'humanité, et non comme un agent normatif autonome. Le risque d'une dynamique technique autoréférentielle échappant progressivement à toute régulation humaine a été souligné (Ellul, 1954/1964). De même, le principe de précaution (Jonas, 1984) rappelle que toute puissance technologique doit s'accompagner d'une responsabilité proportionnée. Par ailleurs, il a été avancé que la technique ne saurait se substituer aux pratiques vertueuses, seules capables d'ordonner les moyens vers des fins véritablement humaines (MacIntyre, 1999). Concevoir la technique comme

auxiliaire implique ainsi de replacer la personne au centre de la normativité, plutôt que la simple performance des artefacts.

Dans le même esprit, une éthique cohérente de l'intelligence artificielle doit être intégrée à des dispositifs juridiques adaptés à la complexité du phénomène, sans être réduite à un simple problème technique. Une éthique de l'information a été proposée (Floridi, 2013) qui, malgré son intérêt, demeure trop abstraite pour fonder un cadre normatif centré sur la personne. De manière complémentaire, une éthique interculturelle de l'information, attentive aux contextes, a également été développée (Capurro, 2008), mais elle requiert un ancrage plus solide dans l'anthropologie philosophique. En définitive, toute régulation véritablement humaine de l'intelligence artificielle suppose une conception métaphysique du sujet, seule capable d'orienter la formulation des droits, des obligations et des limites. Il ne s'agit pas simplement d'établir un consensus minimal, mais de préserver la primauté de la personne dans un contexte où l'efficacité tend à devenir l'unique critère de légitimité.

Une telle réflexion conduit à repenser la relation entre les êtres humains et les dispositifs intelligents, non dans une logique de concurrence ou de substitution, mais selon une complémentarité ordonnée. Une telle relation ne peut être juste qu'à condition de respecter la primauté ontologique du sujet humain, en évitant à la fois l'idéalisation technologique et la dissolution de l'humain. Le principe de subsidiarité, initialement formulé dans le domaine social (Pius XI, 1931), trouve ici un prolongement particulièrement pertinent : les systèmes artificiels doivent être conçus pour assister et non pour remplacer. Dans cette perspective, il a été plaidé en faveur d'architectures éthiques intégrées renvoyant toujours au jugement humain, afin de garantir que la décision morale demeure la prérogative du sujet (Wallach & Allen, 2009).

Par conséquent, la capacité de jugement moral demeure propre à l'homme. Aucun système algorithmique, même parmi les plus sophistiqués, ne possède la conscience réflexive ni l'intentionnalité requises pour l'exercice de la responsabilité. Des mises en garde ont été formulées contre les dangers d'une délégation de la décision morale aux machines, processus impliquant une profonde dépersonnalisation (Sandel, 2009). De manière complémentaire, il a été rappelé que l'intentionnalité, condition de possibilité de toute moralité, demeure indissociable de la conscience et, par conséquent, inaccessible aux systèmes artificiels (Searle, 1992). Nier un tel lien reviendrait à brouiller la frontière entre agent technique et sujet moral, au prix de graves conséquences éthiques et juridiques.

Il en découle que la gouvernance de l'intelligence artificielle doit être orientée vers l'épanouissement humain et non vers l'expansion autonome de la performance technique. La proposition reliant les politiques publiques au développement des capacités humaines (Sen & Nussbaum, 1993) fournit un cadre particulièrement fécond pour concevoir l'intelligence artificielle comme vecteur de service plutôt que comme instance normative autonome. Une telle orientation est renforcée par l'appel à cultiver des vertus techno-morales adaptées aux environnements numériques (Vallor, 2016). La finalité ne peut être que l'eudaimonia, entendue comme accomplissement intégral de la personne humaine.

Conclusions

L'enquête sur la spécificité ontologique de la personne humaine face aux avancées de l'intelligence artificielle permet de mettre en évidence une distinction catégorielle irréductible entre la subjectivité du vivant et le traitement cybernétique. Il est plausible de soutenir que le paradigme fonctionnaliste, en assimilant la simulation opératoire à la compréhension sémantique, néglige des dimensions substantielles de l'intériorité. Dès lors, la redécouverte d'une anthropologie hylémorphique offre un cadre rigoureux pour montrer que la conscience, la liberté et l'affectivité ne constituent pas des propriétés émergentes de la complexité algorithmique. Au contraire, ces facultés apparaissent comme les manifestations d'une forme constitutive qui unifie et oriente l'être humain à partir de sa nature immatérielle et incarnée.

Sur le plan strictement conceptuel, la convergence entre la métaphysique classique et la phénoménologie favorise le dépassement des réductionnismes matérialistes et informationnels contemporains. En distinguant avec précision le traitement syntaxique, propre aux architectures computationnelles, et l'appréhension sémantique, exclusive à l'intellect humain, se trouve délimité l'espace d'une intentionnalité cognitive authentique. De même, la conceptualisation de la téléologie immanente du vivant, opposée à la finalité hétéronome de l'artefact, fournit une clé analytique décisive. Une telle distinction souligne que la finitude et l'anticipation de la mort structurent l'identité narrative et morale de la personne, tandis que les dispositifs techniques fonctionnent selon une logique d'obsolescence matérielle dépourvue de tout horizon existentiel.

Les prémisses ontologiques ainsi dégagées définissent un ensemble d'implications normatives majeures pour le développement technologique et sa gouvernance. Si la capacité de porter des jugements moraux suppose

une intériorité réflexive et une affectivité vécue, il devient éthiquement et juridiquement problématique de déléguer des décisions critiques à des systèmes automatisés. Par conséquent, l'intégration de l'intelligence artificielle dans la société requiert l'adoption du principe de subsidiarité technique. Une telle approche normative garantit que la conception algorithmique demeure un instrument ordonné au bien commun et à l'épanouissement humain, en empêchant toute symétrisation visant à attribuer personnalité, agentivité ou responsabilité à des entités dépourvues d'autonomie réelle.

À partir de ces analyses, de multiples perspectives s'ouvrent pour les recherches philosophiques et juridiques à venir. Il paraît nécessaire d'approfondir l'élaboration de cadres normatifs intégrant des critères ontologiques stables, afin d'éviter que l'évolution technique ne dicte à elle seule la normativité sociale. De même, l'analyse systématique des vertus techno-morales permettant une coexistence harmonieuse avec les environnements numériques constitue un champ d'étude prioritaire. L'examen des transformations que les technologies émergentes imposent aux structures de la médiation intersubjective représente un défi majeur, appelant un dialogue continu entre l'éthique de l'information, la théorie du droit et l'anthropologie philosophique.

Dans cette perspective, l'expansion de l'intelligence artificielle ne représente pas seulement un défi d'ingénierie ou de contrôle systémique, mais engage également une interrogation fondamentale sur le sens de la vie humaine. Affirmer l'irréductibilité spirituelle de la personne face à la simulation cybernétique constitue dès lors une exigence essentielle de préservation humaniste. Il devient ainsi possible de soutenir que le véritable progrès technoscientifique n'atteint sa pleine légitimité qu'à condition d'être subordonné à l'eudaimonia, en reconnaissant que la dignité inaliénable du sujet rationnel demeure la limite fondatrice et la finalité ultime de tout ordre institutionnel.

Références

- Agustine. (2001). *Confessions* (H. Chadwick, trad.). Oxford University Press.
- Aristote. (1991). *De Anima* (R. D. Hicks, trad.; 25th illus. ed.). Prometheus Books.
- . (2001). Physics. In J. Barnes (Ed.), *The Complete Works of Aristotle: The Revised Oxford Translation* (R. P. Hardie & R. K. Gaye, trads., Vol. 1, pp. 315-446). Princeton University Press.
- Block, N. (1980). Troubles with Functionalism. In N. Block (Ed.), *Readings in Philosophy of Psychology* (Vol. 1, pp. 268-305). Harvard University Press.

- . (1981). Psychologism and Behaviorism. *Philosophical Review*, 90(1), 5-43. <https://doi.org/10.2307/2184371>
- Boden, M. A. (1990). *The Creative Mind: Myths and Mechanisms*. Basic Books.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
- Brentano, F. (1995). *Psychology from an Empirical Standpoint* (L. L. McAlister, ed.; A. C. Rancurello, D. B. Terrell & L. L. McAlister, trans.). Routledge (original publicado en 1874).
- Capurro, R. (2008). On Floridi's Metaphysical Foundation of Information Ecology. *Ethics and Information Technology*, 10(2-3), 167-173. <https://doi.org/10.1007/s10676-008-9162-x>
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Deacon, T. W. (2011). *Incomplete Nature: How Mind emerged from Matter*. W. W. Norton.
- Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Company.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.
- Dreyfus, H. L. (1992). *What Computers still can't do: A Critique of Artificial Reason*. MIT Press.
- Ellul, J. (1964). *The Technological Society* (J. Wilkinson, trad.). Vintage Books (original publicado en 1954).
- Finnis, J. (1980). *Natural Law and Natural Rights*. Oxford University Press.
- Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.
- Gehlen, A. (1950). *Der Mensch: Seine Natur und seine Stellung in der Welt*. Aula-Verlag.
- Guardini, R. (1950). *The End of the Modern World*. Henry Regnery Company.
- Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, 42(1-3), 335-346. [https://doi.org/10.1016/0167-2789\(90\)90087-6](https://doi.org/10.1016/0167-2789(90)90087-6)
- Heidegger, M. (1962). *Being and Time* (J. Macquarrie & E. Robinson, trans.). Harper & Row (original publicado en 1927).
- Hui, Y. (2016). *The Question concerning Technology in China: An Essay in Cosmotronics*. Urbanomic.
- Husserl, E. (1999). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First book* (F. Kersten, trad.). Kluwer Academic (original publicado en 1913).
- Jonas, H. (1966). *The Phenomenon of Life: Toward a Philosophical Biology*. Harper & Row.

- . (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press.
- Kane, R. (1996). *The Significance of Free Will*. Oxford University Press.
- Kant, I. (1997). *Groundwork of the Metaphysics of Morals* (M. Gregor, trad.). Cambridge University Press (original publicado en 1785).
- . (1998). *Critique of Pure Reason* (P. Guyer & A. W. Wood, trads. & eds.). Cambridge University Press (original publicado en 1781).
- Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-body Problem and Mental Causation*. MIT Press.
- Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly*, 64(4), 354-361. <https://doi.org/10.1111/j.1468-0114.1983.tb00207.x>
- Lonergan, B. J. F. (1957). *Insight: A Study of Human Understanding*. Longmans, Green and Co.
- Lopes, D. M. (2014). *Beyond Art*. Oxford University Press.
- MacIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. University of Notre Dame Press.
- . (1999). *Dependent Rational Animals: Why Human Beings need the Virtues*. Open Court.
- Marcel, G. (1949). *The Mystery of Being: Volume I. Reflection and Mystery* (G. S. Fraser, trad.). St. Augustine's Press.
- . (1951). *The Mystery of Being: Volume II. Faith and Reality* (R. Rosthal, trad.). St. Augustine's Press.
- Maritain, J. (1953). *Creative Intuition in Art and Poetry*. Pantheon Books.
- . (1994). *The Person and the Common Good* (J. J. Fitzgerald, trad.). University of Notre Dame Press (original publicado en 1947).
- Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel.
- McGinn, C. (1999). *The Mysterious Flame: Conscious Minds in a Material World*. Basic Books.
- Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435-450. <https://doi.org/10.2307/2183914>
- Nancy, J.-L. (2000). *Being Singular Plural* (R. D. Richardson & A. E. O'Byrne, trads.). Stanford University Press.
- Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge University Press.
- . (2006). *Frontiers of Justice: Disability, Nationality, Species Membership*. Harvard University Press.

- Oderberg, D. S. (2005). *Real Essentialism*. Routledge.
- . (2007). Teleology: Inorganic and Organic. In A. J. Beckwith (Ed.), *Contemporary Perspectives on Natural Teleology* (pp. 137-164). Cambridge Scholars Publishing.
- Picard, R. W. (1997). *Affective Computing*. MIT Press.
- Pius XI. (1931). *Quadragesimo anno* [Carta encíclica]. https://www.vatican.va/content/pius-xi/la/encyclicals/documents/hf_p-xi_enc_19310515_quadragesimo-anno.html
- Polanyi, M. (1966). *The Tacit Dimension*. University of Chicago Press.
- Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. Oxford University Press.
- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion* (pp. 37-48). University of Pittsburgh Press.
- . (1975). *Mind, Language and Reality* (Vol. 2). Cambridge University Press.
- Ricoeur, P. (1965). *De l'interprétation: Essai sur Freud*. Éditions du Seuil.
- . (1990). *Oneself as Another* (K. Blamey, trad.). University of Chicago Press.
- Ross, J. F. (1992). Immaterial Aspects of Thought. *The Journal of Philosophy*, 89(3), 136-150. <https://doi.org/10.2307/2026790>
- Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.
- Sandel, M. J. (2009). *Justice: What's the Right Thing to Do?* Farrar, Straus and Giroux.
- Scheler, M. (1928). *Die Stellung des Menschen im Kosmos*. Francke.
- . (1973). *Formalism in Ethics and non-formal Ethics of Values* (M. S. Frings & R. L. Funk, trads.). Northwestern University Press (original publicado en 1913-1916).
- Scruton, R. (2009). *Beauty: A very Short Introduction*. Oxford University Press.
- Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417-457. <https://doi.org/10.1017/S0140525X00005756>
- . (1992). *The Rediscovery of the Mind*. MIT Press.
- Sen, A. & Nussbaum, M. C. (Eds.). (1993). *The Quality of Life*. Oxford University Press.
- Simondon, G. (1958). *Du mode d'existence des objets techniques*. Éditions Aubier.
- Spaemann, R. (1989). *Persons: The Difference between Someone and Something* (O. O'Donovan, trad.). Oxford University Press.
- . (1996). *Glück und Wohlwollen: Versuch über Ethik*. Klett-Cotta.
- Stein, E. (1989). *On the Problem of Empathy* (W. Stein, trad.). ICS Publications (original publicado en 1917).

- Taylor, C. (1985). *Human Agency and Language: Philosophical Papers I*. Cambridge University Press.
- . (1989). *Sources of the Self: The Making of the Modern Identity*. Harvard University Press.
- . (1994). The Politics of Recognition. In A. Gutmann (Ed.), *Multiculturalism: Examining the Politics of Recognition* (pp. 25-73). Princeton University Press.
- Tolstói, L. (1995). *What is Art?* (R. Pevear & L. Volokhonsky, trans.). Penguin Books (original publicado en 1897).
- Thomas de Aquinas. (1947). *Summa Theologica* (Fathers of the English Dominican Province, trans.). Benziger Bros.
- Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin*, 215(3), 216-242. <https://doi.org/10.2307/25470707>
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>
- Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future worth Wanting*. Oxford University Press.
- Wallach, W. & Allen, C. (2009). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.



Publicado bajo una Licencia Creative Commons
Atribución-NoComercial 4.0 Internacional