Le bût de l'Intelligence Artificielle
par Mgr Francesco Follo

Le thème de l'Intelligence Artificielle suscite aujourd'hui l'intérêt, tant par l'engagement et la réflexion de nombreux scientifiques au niveau international – et ce dans plusieurs disciplines telles que l'ingénierie, la logique et les mathématiques, les neurosciences et la philosophie- que dans l'imaginaire collectif, comme en témoigne la vaste filmographie sur le sujet.

Cet événement promu par l'UNESCO est un signal clair de l'importance de cette thématique dans notre époque.

Une machine, pourra-t-elle penser, dans la pleine acception du terme ? Pourra-t-elle égaliser complètement un être humain et sa vie intellectuelle intégrale ? Ce sont quelques-unes des interrogations plus profondes et radicales que les développements autour de l'Intelligence Artificielle posent à l'humanité : scientifiques, hommes politiques et sociologues, philosophes et hommes de foi.

L'histoire de l'Intelligence Artificielle, qui commence réellement dans les années 50 du siècle dernier, pourrait être présentée comme une histoire d'alternance entre des moments de grand optimisme et des phases de prise de conscience des difficultés à reproduire l'intelligence "naturelle", celle de l'homme. Ce n'est ici pas le lieu pour parcourir cette histoire. Cependant il est important de souligner que l'évaluation des résultats obtenus dépend d'une manière essentielle des objectifs que la communauté scientifique et technologique se fixe au fur et à mesure. De ce point de vue, importante est la distinction, désormais devenue canonique, entre une Intelligence Artificielle "forte" et une intelligence artificielle "faible".
L'Intelligence Artificielle faible pose pour objectif la conception et la construction de machines qui agissent comme si elles étaient intelligentes. L'Intelligence artificielle forte, quant à elle, se pose comme objectif ultime de projeter des machines effectivement similaires à l'homme, jusqu'au point de pouvoir développer une conscience de soi.
A cette distinction s'en ajoute une autre : celle selon laquelle obtenir une intelligence artificielle impliquerait une reproduction soit complète du substrat matériel qui permet à l'homme d'avoir une démarche rationnelle - le cerveau- soit simplement du résultat (ou de certains des résultats) de cette même pratique rationnelle indépendamment des spécificités physiques ou d'ingénierie élaborées pour la structure de la "machine pensante".

Aujourd'hui, une machine reproduisant les détails de l'organisation cérébrale et dotée de toutes les caractéristiques les plus élevées de l'intelligence humaine, y compris la conscience de soi, semblerait aller au delà de la portée des développements techniques actuels ou prévisibles. Cela dépend aussi du fait que, malgré les importants progrès des neurosciences, nous ne disposons pas aujourd'hui d'une image claire et complète de comment est constitué et fonctionne le cerveau. Cependant il existe des machines (ou des programmes informatiques) qui sont capables d'accomplir des opérations complexes avec des prestations comparables, ou parfois supérieures, à celles des hommes, et qui ont déjà quitté le terrain des laboratoires de recherches en se diffusant dans plusieurs applications techniques de niche ou sont même disponibles dans le commerce. Cet aspect alimente un optimisme diffusé qui laisse ouverte pour certains la possibilité, même lointaine, de satisfaire un jour aux conditions exigées par l'Intelligence Artificielle forte.

Les courtes notes qui suivent se posent l'objectif de réfléchir sur certains aspects qui, à mon avis, pourraient été destinés à déterminer les différences qualitatives ou "essentielles" qui subsistent entre Intelligence Artificielle et intelligence humaine. Il s'agit de réflexions philosophiques qui prennent en compte certains des aspects techniques et scientifiques concernés.

Le problème apparaît beaucoup moins profond si on prend en considération l'Intelligence Artificielle faible, car, par définition, l'objectif est dans ce cas de produire des machines capables d'agir comme si elles étaient intelligentes sans prétendre être de réelles reproductions de l'intelligence humaine. Aussi, les résultats déjà obtenus- qui sont parfois réellement surprenants- concernent des situations et des compétences spécifiques et assez réduites. Cependant, si les développements futurs permettaient en effet de produire des machines capables d'agir dans chaque situation ou contexte comme si elles étaient intelligentes, et dans chaque domaine de connaissance et d'action propre à l'homme, cela nous rassurerait dans l'un des aspects de l'intelligence artificielle forte, c'est-à-dire la possibilité de reproduire *in toto*, l'intelligence humaine. En effet : dans un cas pareil, on ne verrait pas encore réalisé le rêve de reproduire non seulement les prestations mais également la structure ou la constitution du substrat de l'intelligence humaine. Cependant si nous considérons que, comme il a été dit, nous n'avons pas encore aujourd'hui une image claire et complète de la manière dont notre cerveau est constitué et fonctionne, la réalisation de machines capables d'agir comme si elles étaient intelligentes, dans chaque situation et dans chaque domaine, poserait dans tous les cas et de manière radicale le problème de l'unicité et de la non reproductibilité de l'intelligence humaine.

Il existe, à mon avis, au moins trois aspects qui rendent difficile de penser qu'un jour les machines pourront se substituer entièrement à l'intelligence humaine : la dimension affective (ou émotionnelle), la dimension sémantique et une troisième dimension que pour l'instant j'appellerai « motivationnelle ».

La tradition philosophique a toujours reconnu, d'une manière ou d'une autre, que les attaches, les émotions, les sentiments et les appétits influencent la cognition humaine. Aujourd'hui il paraît clair, même d'un point de vue expérimental, que les états émotionnels influencent les processus décisionnels et rationnels : il s'agit des choix. Et cela, indépendamment du thème de l'Intelligence Artificielle, pose des problèmes vis-à-vis de la « rational choice theory ». Suivant cette approche, une décision rationnelle suivrait un processus entièrement logique qui cherche à obtenir une décision optimale qui maximise l'utilité et minimise les risques. Les processus décisionnels réels de l'homme ne suivent pas toujours cette même procédure, aussi, mais pas seulement, pour l'importance des états émotionnel comme souligné par ce qu'on appelle l'« affect heuristics ». Les objectifs non plus ne paraissent pas être toujours la maximalisation des résultats, car les décideurs réels peuvent souvent se contenter de solutions suffisamment bonnes même si elles ne sont pas optimales (voir pour cela la différence entre « maximizers » et « satisficiers »). Si les choses sont ainsi, une Intelligence Artificielle forte qui voudrait reproduire entièrement l'intelligence humaine, devrait aussi reproduire les aspects émotionnels et affectifs.

Or, premièrement, aujourd'hui il ne paraît pas possible reproduire de manière artificielle la sphère émotionnelle et sentimentale qui caractérise de manière diffuse l'intelligence humaine : et ce fait constitue une limite assez importante au rêve de l'Intelligence Artificielle forte. Deuxièmement, le fait d'introduire la dimension émotionnelle dans des «machines lourdes » rendrait leur fonctionnement moins rationnel (dans le sens de la « rational choice theory »). De manière

collatérale, ces considérations posent des interrogations même à ceux qui s'occupent d'Intelligence Artificielle d'un point de vue technique : quel est l'objectif de ces développements technologiques ? Reproduire et imiter l'intelligence humaine ou soutenir cette dernière dans des situations et domaines spécifiques en laissant les aspects intégraux et compréhensifs à l'homme fait de chair et d'os, de raison et de sentiments ?

A ce propos, j'aimerais ensuite indiquer que la tradition philosophique (et théologique aussi) a toujours souligné l'importance de l'amour et de la charité pour la recherche même de la vérité. La poursuite d'objectifs cognitifs élevés demande une tension vers la vérité qui ne peut pas être expliquée complètement en termes d'utilité ou d'intérêt. C'est un aspect de cette dimension motivationnelle dont nous parlions toute à l'heure et sur laquelle je reviendrai par la suite. Pour l'instant, il vaut la peine d'indiquer qu'une machine intelligente voulant imiter et reproduire totalement l'intelligence humaine devrait prendre en considération cet aspect que je qualifierais d'"émotionnel élevé", en rapport naturellement à ce qui a déjà été dis concernant l'influence générale qu'ont les états émotionnels sur les processus cognitifs et décisionnels.

Un deuxième aspect problématique de l'Intelligence Artificielle (forte) concerne la distinction - assez classique dans la philosophie moderne et contemporaine du langage- entre syntaxe et sémantique. Assez diffusée est l'idée que les calculateurs, y compris ceux qui devraient soutenir une intelligence artificielle, puissent analyser seulement les liens syntaxiques entre symboles non interprétés (c'est à dire sans signification), alors qu'ils ne pourraient pas traiter les contenus sémantiques qu'il est possible d'attribuer à ces symboles. Ce fait poserait une énorme limite à la possibilité de réaliser une Intelligence Artificielle qui puisse imiter *in toto* l'intelligence humaine, tant cette dernière est imprégnée de significations à plusieurs niveaux. A ce propos, l'idée est souvent proposée qu'en réalité même la sémantique de haut niveau pourrait être réduite à la syntaxe, et que la limite serait seulement celle de la puissance de calcul et de la complexité requise aux structures physiques artificielles afin de gérer une telle sémantique "syntaxisée" ; qu'ainsi dans le futur on pourra avoir des machines capables de traiter même les aspects sémantiques les plus sophistiqués montrés par l'intelligence humaine. A ce sujet je crois que deux considérations pourraient être appropriées et intéressantes.

Dans un premier temps, en prenant les distances des approches classiques de la linguistique – qui voient syntaxe et sémantique comme deux aspects différents du langage, la syntaxe définirait une série de règles générales de composition applicables dans des vastes catégories d'éléments linguistiques, sans égard à la signification qui en résulterait - il existe aujourd'hui des approches en linguistique – de plus en plus liées à des développements neuroscientifiques- qui proposent une plus grande interdépendance entre sémantique et syntaxe (par exemple, la « Cognitive grammar » et la « Construction grammar »). Le point le plus important est que ces développements suggéreraient une « réduction » de la syntaxe à la sémantique plutôt que l'inverse ! En d'autres termes, les constructions syntaxiques utilisées dans le langage et dans le raisonnement humain seraient intrinsèquement dépendantes des concepts et/ou significations des expressions « combinées » - c'est-à-dire de la sémantique- et qu'il n'existerait donc pas une « syntaxe séparée » généralement et génériquement applicable aux éléments signifiants du langage (ou de la pensée : les concepts).

Dans un deuxième temps, il ne faudrait pas perdre de vue ce qu'est la sémantique. On pense souvent que la sémantique est le réseau de relations entre mots. Par exemple, si nous recherchons dans un

dictionnaire un mot, ce mot est défini par d'autres mots liés. « Calendrier » est défini comme un ensemble de feuilles qui marquent les jours, les semaines, les mois dans l'année. Celui qui connaît la signification de tous les mots de cette définition pourrait comprendre le mot « calendrier » ; celui qui n'en connaît aucun, pourrait continuer à rechercher sur le dictionnaire et cela de manière récursive.

Cependant, même en continuant dans ce processus de recherche sur le dictionnaire, l'individu qui n'avait jamais vu ou possédé un calendrier, aurait difficilement pu avoir une compréhension authentique de ce mot. Il est clair ensuite que la compréhension d'un mot dépend aussi de manière profonde d'expériences réelles faites par l'individu. La signification de mots comme « pauvreté » ou « liberté » varie selon la situation personnelle, sa propre histoire, et du contexte historique et géographique. Les significations sont connotées d'une manière aussi bien émotionnelle que rationnelle.

Sur ces bases, pour avoir la réalisation de machines pouvant – si l'objectif est l'Intelligence Artificielle forte- reproduire chaque aspect de l'intelligence humaine, il ne faudrait pas seulement qu'elles soient dotées d'une puissance suffisante de calcul mais aussi, en emphatisant, qu'elles vivent comme un être humain. Qu'elles n'analysent pas seulement les symboles mais qu'elles fassent également des expériences, qu'elles souffrent, se réjouissent, désirent, aient peur, voient, entendent, touchent, sentent et goutent. L'Intelligence Artificielle impliquerait donc une « Vie artificielle » et comme vous le savez, les obstacles rencontrés dans les tentatives de produire des machines « vivantes » sont tant nombreuses et complexes, que ceux rencontrés par l'Intelligence Artificielle.

J'aimerais ainsi aborder le troisième aspect que j'avais prévu d'évoquer, celui que j'ai appelé l'aspect « motivationnel ». Il est clair, qu'aujourd'hui, plusieurs aspects de l'intelligence humaine découlent du long parcours évolutif qui a amené à notre espèce biologique. Certains de ces aspects sont même partagés, au moins partiellement, avec d'autres espèces animales non humaines. Il est aussi connu que selon les théories concernant l'évolution biologique, les nouveautés qui ont émergé dans l'histoire naturelle, même en ce qui concerne le comportement et la cognition, répondent à une logique de nécessité. De nécessité pas dans le sens que le processus évolutif serait d'une certaine manière nécessaire, mais dans le sens que pendant l'histoire de l'évolution il émerge, pour les différentes espèces biologiques dans les différents contextes environnementaux, ce qui est nécessaire pour survivre – sans quoi on succomberait aux défis environnementaux.

Ce cadre conceptuel, même valable pour plusieurs aspects de la connaissance humaine, semble ne pas être capable d'expliquer entièrement ce qu'on appelle l'évolution culturelle qui caractérise au moins les derniers 15-20 millénaires de l'histoire humaine.

La domestication de plantes et d'animaux, la construction de villes et de lieux de culte, l'invention de l'écriture et de l'arithmétique, la naissance de ce qu'on appelle la culture théorique, des universités, de la science moderne, les révolutions industrielles qui ont marqué les derniers siècles : tout cela est difficilement dépendant uniquement de la nécessité dont nous avons parlé tout à l'heure.

Aucune de ces innovations (et les nombreuses autres inventions spécifiques qui les ont accompagnées) n'était stricto sensu nécessaires à la survie de l'être humain. La question résulte encore plus profonde si regardée du point de vue des individus uniques qui ont donné des contributions essentielles à ces avancements. Pensons à Socrate, qui, par amour de la vérité et de la justice a été obligé de se suicider ; ou à Galilée qui, parce qu'il soutenait ses convictions

cosmologiques a dû affronter deux procédures judiciaires. Ces deux réflexions sommaires – auxquelles pourrait s'ajouter d'autres exemples - posent avec force le problème des motivations qui poussent l'être humain à rechercher, à vouloir connaître et à inventer – qui le poussent vers la vérité et à vouloir améliorer ses propres conditions matérielles et spirituelles au delà de ce qui est strictement nécessaire.

Il s'agit d'une thématique vaste, qui ne peut pas être épuisé à cette occasion. Cependant elle pose un autre problème par rapport à l'objectif de l'Intelligence Artificielle forte. Une machine pensante voulant reproduire *in toto* l'intelligence humaine devrait aussi savoir reproduire cet aspect motivationnel. En d'autres termes, elle ne devrait pas seulement être capable d'opérations intelligentes pour accomplir des taches attribuées par d'autres, mais devrait être aussi capable de s'auto-attribuer des taches et des objectifs, et d'avoir des aspirations. Aujourd'hui, des réseaux de neurones sophistiqués et ce qu'on appelle les « systèmes experts » sont capables de conduire des opérations intelligentes de manière comparable - et des fois même supérieure - à l'être humain, et peuvent le faire même en essayant de trouver des solutions non pré-insérées dans le système (on peut penser à ce que l'on appelle « algorithmes génétiques », c'est-à-dire des programmes capables de se modifier de manière autonome afin de modifier ses propres prestations). Cependant, non seulement elles ne peuvent faire ça que dans des domaines limités et circonscrits mais surtout elles ne s'auto-attribuent pas des objectifs à réaliser. Une réelle Intelligence Artificielle forte devrait en revanche être capable de le faire. Ce n'est pas par hasard si plusieurs films sur le sujet proposent le thème de la machine qui se rebelle, qui veut s'autodéterminer ou qui souhaite devenir humaine, ou qui veut protéger l'humanité plutôt que la soumettre et conquérir la Terre. Il me semble qu'aucun des développements en cours ou réellement prévisibles permettrait d'arriver à un tel résultat.

En conclusion, je voudrais livrer deux autres brèves réflexions. La première est qu'évidemment les trois aspects que j'ai voulu exposer sont étroitement liés et même si je considère que le dernier soit le plus profond, il parait clair que les motivations sont étroitement liées tant à la dimension affective qu'à la dimension sémantique. La deuxième réflexion conclusive a l'intention de toucher un thème philosophique et théologique central dans la tradition chrétienne : la question de l'âme humaine. Je n'ai pas l'intention ici d'aborder la question dans toute son amplitude, mais seulement de souligner que le thème de la motivation qui pousse l'homme à s'autodéterminer et à s'auto-surpasser, même au-delà des nécessités, a des liens assez étroits avec certaines des « fonctions » que la tradition philosophique et théologique chrétienne attribue à l'âme : celles concernant la liberté, la dignité, l'auto-conscience et la conscience morale. Il est donc à noter que certains des problèmes qui émergent des objectifs – et des interrogations - imposés par l'Intelligence Artificielle, pourraient trouver des équivalents, des résonances et des raisons d'approfondissement même dans certaines thématiques philosophiques-théologiques traditionnelles à caractère éminemment anthropologiques.