

L'IA peut-elle démultiplier l'Homme sans nuire ?

Par Jacques PIGNAULT

Expert en maîtrise des risques et facteur humain

Doté de la pensée réfléchie qui le caractérise, poussé par son instinct de survie et de domination de son milieu, l'Homme a toujours utilisé son génie inventif et sa faculté de réflexion à développer les outils qui lui semblaient de nature à prolonger ses capacités physiques et intellectuelles :

- ❖ Mieux faire (plus vite, moins cher, plus adapté) ce qu'il peut faire,
- ❖ Faire ce qu'il ne sait (ou ne peut) pas faire

Au fil des âges, plus sophistiqués étaient ses outils, plus grandes étaient ses possibilités, et plus ambitieux étaient les renforcements dont l'Homme se dotait en matière d'acquisition d'informations, de réflexion et d'action (cf. figure 1).

Activités	Limitations humaines	Renforcements escomptés
Acquisition	<ul style="list-style-type: none"> ❖ La limitation des capteurs à cinq sens ❖ L'acquisitions d'un faible nombre de données ❖ Le biais psychologique de la capture (préférences) ❖ La versatilité et l'inconstance des humains 	Augmenter : <ul style="list-style-type: none"> ❖ Les moyens de capture ❖ Le nombre d'acquisitions ❖ L'impartialité et la constance ❖ La capacité de mémorisation
Réflexion	La réflexion repose essentiellement sur des bases individuelles : <ul style="list-style-type: none"> ❖ Les expériences passées ❖ Les connaissances (formation) ❖ La culture, les croyances, l'idéologie ❖ La forme physique et psychologique 	<ul style="list-style-type: none"> ❖ Elargir les bases de la réflexion ❖ S'affranchir des idées préconçues et des « dictats » ❖ Réduire voire supprimer les aspects subjectifs
Action	Difficulté de modifier une image mentale établie : <ul style="list-style-type: none"> ❖ La focalisation sur l'attendu, ❖ La faible attention sur les évolutions non escomptées 	<ul style="list-style-type: none"> ❖ Faciliter les remises en cause justifiées ❖ S'affranchir des contraintes physiques et psychologiques (stress notamment) ❖ Assister le contrôle et le pilotage ❖ Associer aux choix qui s'offrent différentes conséquences possibles

Figure 1 : Dépasser nos limites physiques et intellectuelles de base [5]

Cette évolution progressive nous a conduits à un état dans lequel le couplage des technologies de l'information et des systèmes robotisés ouvre la porte à un changement profond dans l'organisation de nos activités et de nos modes de réflexion et de pensée. Cette modification touche la plupart des aspects de notre vie, tels l'habitat, le transport, la médecine, le sport, et les mondes militaire, scientifique, industriel, commercial et administratif.

Il faut bien souligner que le domaine privilégié des applications de l'Intelligence Artificielle (IA) est la résolution de problèmes à solutions probabilistes (solutions vraisemblables à des niveaux de confiance variables, la météo par exemple). Les résultats obtenus correspondent alors aux solutions les plus vraisemblables et une interprétation à l'aune des valeurs et du sens est nécessaire. En revanche, lorsque la résolution est de type déterministe (problèmes à solution unique, automatismes de sécurité d'une installation industrielle par exemple), point n'est besoin d'IA, de logique floue et d'apprentissage profond. Ils restent du ressort des algorithmiques « classiques » (par exemple, si la pression excède un seuil donné, ouvrir la soupape de sécurité).

Sur les plans économique et scientifique, les effets des développements réalisés, en cours ou à venir sont très positifs. Il est clair néanmoins que se dressent de redoutables écueils à consonance humaine. Des écueils et non des barrages car ils sont contournables. Ils concernent nos vies :

- ❖ *professionnelles* (pertes d'emplois ou emplois sous contrôle d'un robot, suppression du dialogue et des échanges entre pairs, entre métiers, entre collaborateurs d'une même entreprise) ;
- ❖ *personnelles* (pertes d'autonomie car assistance totale dans la vie courante, pertes des libertés par suivi permanent, délaissement du lien social réel pour le lien social virtuel) ;
- ❖ *familiales* (transmission des valeurs, solidarités familiales) ;
- ❖ *spirituelles* (perte de la place et du projet de l'homme dans l'univers) .

Ces écueils ne doivent cependant pas étouffer les moteurs des travaux en cours qui, dans une vision humaniste, sont de compléter, démultiplier et libérer l'Homme pour lui permettre de focaliser ses efforts sur la recherche et la découverte.

Se pose ainsi le problème de concevoir et d'utiliser au mieux les outils IA pour en tirer la quintessence et en contrôler les effets pervers.

L'examen du processus de création et d'utilisation d'un outil IA nous permet de mettre en exergue les points de vigilance à exercer. De façon très schématique, ce processus se déroule comme présenté en figure 2.

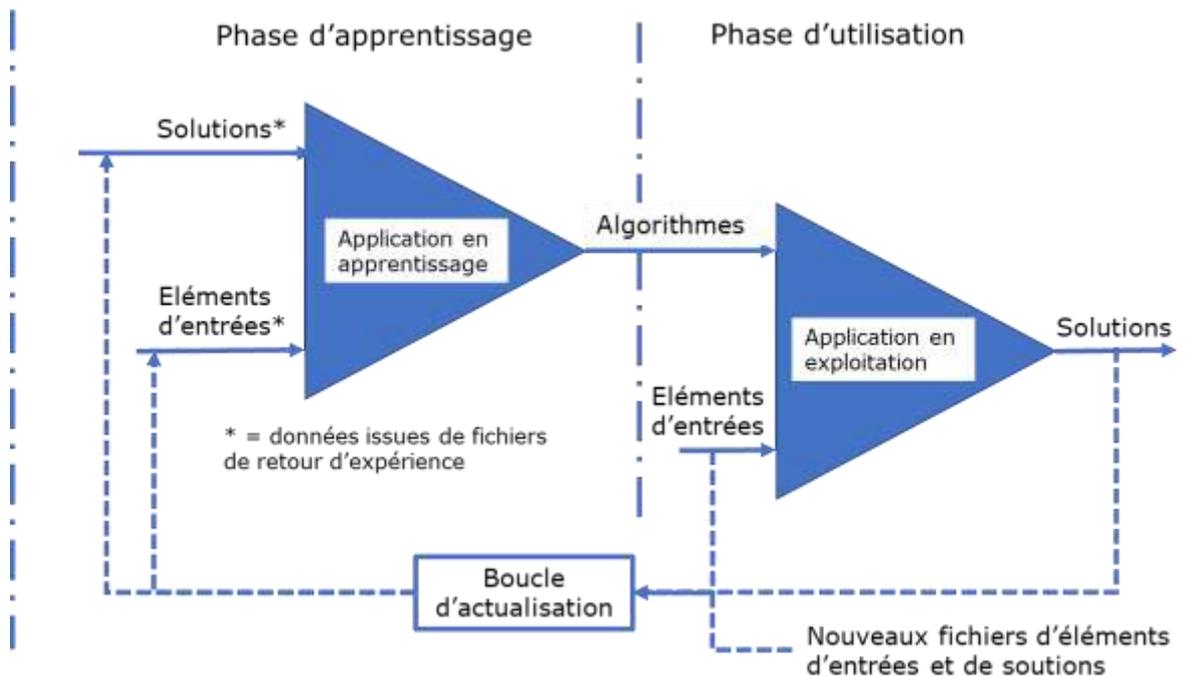


Figure 2 : Phases de vie d'une application IA

La phase d'apprentissage, souvent dénommée « deep learning », a pour objectif la mise au point des *algorithmes* permettant de relier des *solutions* correctes répondant à des problèmes que posent des *éléments d'entrées*, par exemple :

- ❖ à une photo ou à une image graphique en entrée, obtenir la désignation de ce qu'elle représente en sortie,
- ❖ à un jeu de données météo, obtenir les prévisions d'ensoleillement, de vent et de précipitations en sortie,
- ❖ à l'aspect de surface d'un ouvrage d'art (barrage hydroélectrique par exemple), obtenir les actions de maintenance à engager.

Les fichiers d'apprentissage regroupent des éléments d'entrée et les solutions connues qui en résultent. Issus du retour d'expérience ces fichiers doivent être représentatifs des situations envisageables par les futurs utilisateurs de façon aussi exhaustive et pertinente que possible. Cela suppose que les équipes de conception et de réalisation ont un environnement, une culture, une éthique, des croyances et des valeurs morales analogues (sinon identiques) à celles des exploitants futurs.

En exploitation, les applications ne vérifieront en effet que la conformité logique des groupes d'entrées et de sorties, ainsi que leur cohérence d'ensemble avec les algorithmes et données mises en œuvre lors des phases d'apprentissage. Le sens (l'éthique, notions de bien et de mal) des données de sortie (solutions) qu'elles produisent et les valeurs auxquelles les rattacher seront toujours des aspects totalement étrangers aux applications car seul l'humain connaît le sens des actions qu'il engage. Ce sens lui est révélé explicitement (il en est conscient) ou implicitement (il en a seulement l'intuition ou le sentiment, mais ne peut l'expliquer) en application des valeurs auxquelles il croit.

La phase d'exploitation consiste à intégrer dans la vie courante les applications IA pour qu'elles nous renforcent en nous positionnant progressivement sur une trajectoire couramment dénommée « l'Homme augmenté » :

- ❖ *Libérer* l'Homme, et dans un premier temps, s'affranchir des tâches routinières et sans intérêt. Cette liberté nous donnera la possibilité de prendre du recul et de nous axer sur des activités plus humaines.
- ❖ *Démultiplier* l'Homme, pour qu'il soit plus performant, pour qu'il ait une puissance accrue. Cela se traduira par exemple par plus d'activités, plus vite, mieux, de façon plus pertinente.
- ❖ *Compléter* l'Homme, c'est-à-dire accroître la connaissance et la capacité à explorer notre milieu, rendre possible ce qui ne l'était pas du fait de nos limitations humaines naturelles.

Cela signifie une augmentation d'efficacité et de cohérence des actions individuelles et collectives, et non pas un repli de l'Homme sur lui-même par accroissement d'autonomie.

Dans cet esprit, il nous resterait idéalement à développer les thèmes proprement humains :

- ❖ Relations humaines, esprits d'équipe et d'entreprise, mimétique, esthétique ;
- ❖ Affectivité, valeurs morales, notions de bien et de mal, dévouement, sacrifice ;
- ❖ Humour, imagination, intuition, téléologie, prudence, instinct de survie ;
- ❖ Feinte, mensonge, compréhension, émotions (plaindre, féliciter, douter), confiance ;

L'observation et les analyses de type utilisateur réalisées auprès des différents secteurs d'activités concernés par l'IA mettent en exergue des risques de déviations majeures vis-à-vis de la trajectoire « Homme augmenté » recherchée. Elles ont pour origine le développement des applications IA ou (et) leur utilisation.

Les grandes catégories de déviations conceptuelles sont les suivantes :

- ❖ *Le biais* : les algorithmes et l'apprentissage des outils créés ne respectent pas, lorsqu'ils sont déployés, les règles juridiques et éthiques en vigueur, par exemples, les règles encadrant la discrimination de race ou de sexe, limitant l'intrusion dans l'intimité et la vie privée des personnes, contrôlant la création ou le renforcement de partis pris injustifiés.
- ❖ *Le détournement* : les outils créés doivent pouvoir réaliser tout ce qu'on attend d'eux à la conception, et seulement cela. Il est en effet important qu'ils ne puissent pas être utilisés pour des usages autres que ceux prévus (notamment pour nuire au lieu d'aider, pour être mis en œuvre à des fins terroristes ou extrémistes). En utilisation « normale », les outils créés doivent être robustes et bénéfiques pour l'humanité.
- ❖ *La disponibilité* : elle doit être sans faille car leur disponibilité opérationnelle est de plus en plus indispensable à un bon fonctionnement des sociétés humaines (en voiture, ne sommes-nous pas « un peu perdus » -dans tous les sens du terme- lorsque notre GPS est en panne ?).
- ❖ *L'opacité* : une application IA comprend des algorithmes qui traitent des données et élaborent des résultats dans le respect des contraintes. Ces résultats sont proposés à l'utilisateur qui les valide ou non. Ces validations

peuvent être très engageantes, voire vitales (permis de tuer par exemple). Il est donc important que l'utilisateur comprenne le cheminement suivi par l'outil pour aboutir à telle ou telle conclusion. Et ce en temps réel, car certaines situations ne peuvent rester en attente de décision (avionique par exemple). A défaut, l'utilisateur ne peut pas avoir confiance et donc ne peut pas valider. Cette confiance sera par ailleurs nécessaire pour passer des outils d'assistance aux outils autonomes.

L'intégration des applications IA (qu'elles soient ou non porteuses de déviations conceptuelles) dans des activités déplace le rôle initialement dévolu à l'Homme.

L'IA prend en effet progressivement en charge les activités traditionnellement réalisées par l'Homme pour lesquelles les compétences techniques sont substituables : activités rationalisables et codifiables, par exemple et de façon majoritaire, les activités routinières et répétitives.

En revanche, les activités nécessitant des compétences englobant une dimension affective, relationnelle, mimétique ou (et) téléologique ou (et) des capacités d'analyse générale (flexibilité, adaptabilité, aptitude à la résolution de problèmes, créativité) semblent à ce jour non (ou peu) substituables. Il faut donc prioritairement former les hommes dans le domaine des compétences non (ou peu) substituables.

Bien assumer cette évolution nécessite une préparation comportementale et professionnelle qui n'est pas toujours bien ressentie par les personnes impliquées. Il s'ensuit l'ajout des principaux risques suivants :

- ❖ *Postes de travail supprimés* : le remplacement pur et simple de l'homme par le robot est parfois préféré à la substitution de l'homme par un couple homme-robot. Ceci pose des problèmes d'emploi d'une part, et de gestion des connaissances d'autre part. En effet, perdant le savoir-faire car il ne fait plus, l'homme n'est progressivement plus à même d'intégrer les enseignements de type retour d'expérience dans « l'intelligence » du robot.
- ❖ *Intervenants déresponsabilisés* : les résultats obtenus par les applications IA ont une influence importante sur les individus car issus d'algorithmes porteurs de savoir-faire collectif (acquis lors des phases d'apprentissage). L'homme peut alors se sentir en état d'infériorité et pratiquer le « ce n'est pas moi, c'est l'ordinateur qui m'a dit de faire cela ». Du refus de responsabilité au laisser-aller, il n'y a qu'un pas que certains peuvent franchir sans avoir conscience que cette démission ponctuelle de pensée est un cheval de Troie dans la globalité de leur système de pensée consciente et inconsciente.
- ❖ *Libertés individuelles réduites* : il s'agit bien entendu d'abord de nos modes de vie individuels et sociaux, mais aussi de l'évolution de nos pensées vers un formatage unique (celui du collectif). On a coutume de dire que la diversité est une force car elle conduit à des situations de challenge génératrices de progrès. Les algorithmes IA résultant d'une compilation de données collectives, ils font nécessairement abstraction des expertises individuelles non formalisées. L'innovation, et donc le progrès, ne peuvent qu'en pâtir par une forme d'appauvrissement.
- ❖ *Clivage de la société* : intégrer dans son univers mental les technologies de l'information n'est pas immédiat pour tous. En tout état de cause cela nécessite

des connaissances et des évolutions comportementales dont la dynamique d'acquisition et d'assimilation est moins rapide que celle de l'évolution des technologies. En l'état, l'IA tend à créer une société à deux vitesses : les « hommes plus », qui conçoivent et gèrent les applications IA et les « hommes moins » qui seront étrangers au mode dans lequel ils vivent. Seules l'éducation et la volonté des hommes peuvent permettre d'éviter cette régression notable d'humanisme.

Les applications IA sont rapidement et massivement plébiscitées et mises en services dans notre monde d'aujourd'hui. Elles constituent un facteur de progrès indéniable pour notre civilisation.

Moyennant la maîtrise des déviations identifiées, qui chacune entrainerait une régression de l'homme dans sa vie personnelle, familiale, sociale et spirituelle, quel sera l'impact sur notre pensée créative et inventive ? Plus largement l'IA favorisera-t-elle une évolution de la noosphère, telle que définie par le Père Pierre Teilhard de Chardin [7], vers une socialisation totalitaire ou vers une socialisation de communion ?

De prime abord, et si nous ne sommes pas assez attentifs, l'évolution de la noosphère vers une socialisation totalitaire semble inévitable car résultante de la substitution progressive des pensées des individus (diversité) aux applications IA qui reflètent le tronc commun de la pensée collective (uniformité).

Si les hommes sont capables de dominer l'IA qu'ils ont eux-mêmes créée, ils vont avoir plus de temps et de latitude de liberté, car dégagés progressivement des contingences matérielles. S'ils ne laissent pas se développer la tendance « la machine travaille pour moi, j'en profite et je lézarde », ils pourront alors utiliser cette liberté pour penser (cf. figure 3) et aller vers cette délicate socialisation de communion qui allie l'individualité et la pluralité.

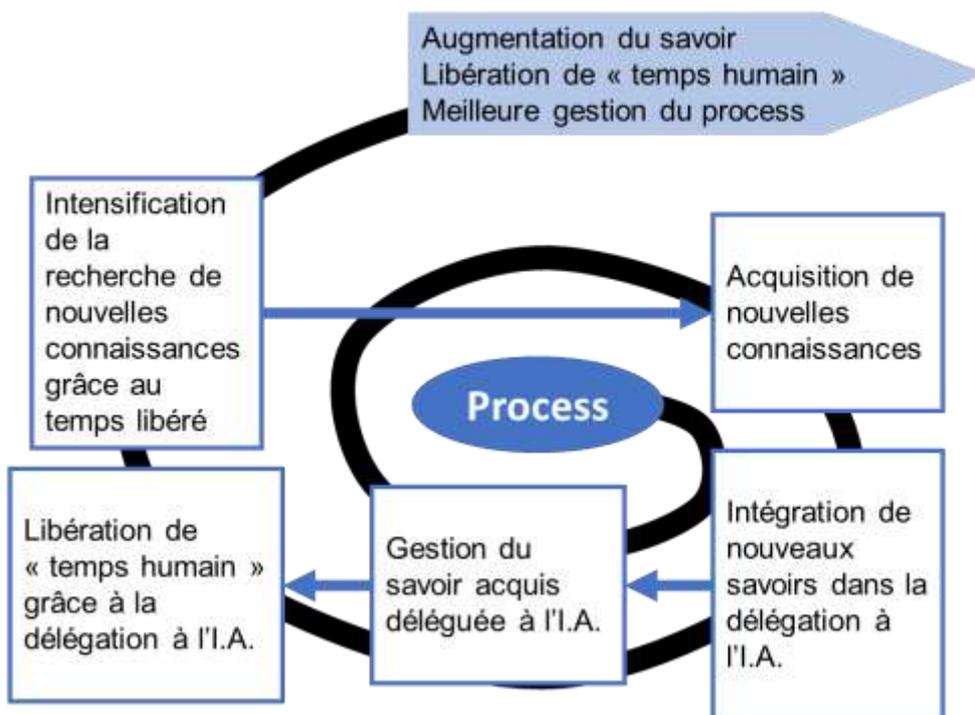


Figure 3 : La spirale de progrès [5]

Dans cette perspective, le « temps humain » libre peut utilement être employé à acquérir de nouvelles connaissances permettant d'accroître le domaine de savoir et donc la délégation à l'intelligence artificielle, et par conséquent de libérer encore plus de « temps humain ». Dans cette spirale, l'homme est de plus en plus affecté à des tâches non substituables de recherche de connaissance, et l'IA de plus en plus chargée de mettre en œuvre les connaissances acquises, sous supervision humaine.

Concilier complexité, diversité et pluralité n'est pas simple. Si nous savons en éviter les écueils, l'IA nous offre une chance d'aller plus vite vers la socialisation de communion en libérant du temps que nous sommes à même de lui consacrer.

En effet, si l'on adopte la représentation systémique [2] de la noosphère précisée par Gérard Donnadiou, à savoir qu'elle comprend les systèmes artificiels, les systèmes sociaux et les systèmes symboliques, il est clair que l'impact du déploiements des applications IA sera de :

- ❖ Faciliter la constitution des systèmes artificiels et sociaux
- ❖ Nous donner plus de temps de pensée pour constituer le système symbolique et évoluer vers la formation de la super intelligence Ω [3] prônée par le Docteur Lui Feng

L'IA est très probablement une chance que nous donnent les technologies de l'information pour faciliter l'émergence de la prophétie du Père Pierre Teilhard de Chardin :

« Quelque jour, après l'éther, les vents, les marées, la gravitation, nous capterons, pour Dieu, les énergies de l'amour. Et alors, une deuxième fois dans l'histoire du monde, l'homme aura trouvé le feu » [8] Pékin , février 1934.

Sources :

1. DESCOUPS Christian, *Liberté et intelligences artificielles : rivales ou alliées ? (colloque tenu à Luxembourg le 18 octobre 2019 à l'initiative de PETRUSSE ASBL)*, mensuel AGEFI Luxembourg, novembre 2019
2. DONNADIEU Gérard, *De la biosphère à la noosphère*, Revue Noosphère, mars 2019
3. Docteur LUI FENG, *La formation de la super intelligence Ω* , Revue Noosphère mars 2020
4. MAGNE Jean, *Ethique et intelligence artificielle*, Colloque « les entretiens du risque » organisé par l'Institut pour la Maîtrise des Risques (IMdR) décembre 2019
5. PIGNAULT Jacques, MAGNE Jean, FOY Bertrand, *Fascinante IA*, Boleine, 2019
6. PIGNAULT Jacques, *L'intelligence artificielle : s'en servir sans s'asservir !*, Revue Les Plumes de l'ARA, décembre 2019
7. TEILHARD de CHARDIN Pierre, *La place de l'homme dans la nature*, édition du seuil, 1956
8. TEILHARD de CHARDIN Pierre, *Les directions de l'avenir*, édition du seuil, 1973