



Le Saint-Siège

LE PAPE FRANÇOIS PARTICIPE A LA SESSION DU G7 SUR L'INTELLIGENCE ARTIFICIELLE
[13-15 juin 2024]

DISCOURS DU PAPE FRANÇOIS

Borgo Egnazia (Pouilles, Italie)
Vendredi 14 juin 2024

[Multimédia]

Un outil fascinant et redoutable

Mesdames et Messieurs !

Je m'adresse à vous aujourd'hui, dirigeants du Forum intergouvernemental du G7, pour vous présenter une réflexion sur les effets de l'intelligence artificielle sur l'avenir de l'humanité.

« L'Écriture Sainte témoigne que Dieu a donné aux hommes son Esprit pour qu'ils aient "la sagesse, l'intelligence et la connaissance de toutes sortes de travaux" (Ex 35, 31) » [\[1\]](#). La science et la technologie sont donc les produits extraordinaires du potentiel créatif des êtres humains [\[2\]](#).

Or c'est précisément l'utilisation de ce potentiel créatif donné par Dieu qui est à l'origine de l'intelligence artificielle.

Cette dernière, comme on le sait, est un outil extrêmement puissant, utilisé dans de nombreux domaines de l'activité humaine : de la médecine au monde du travail, de la culture à la communication, de l'éducation à la politique. Et l'on peut désormais supposer que son utilisation influencera de plus en plus notre mode de vie, nos relations sociales et même, à l'avenir, la manière dont nous concevons notre identité en tant qu'êtres humains [\[3\]](#).

Le thème de l'intelligence artificielle est cependant souvent perçu comme ambivalent : d'une part, il enthousiasme par les possibilités qu'il offre, d'autre part, il suscite la crainte par les conséquences qu'il laisse présager. À cet égard, on peut dire que nous sommes tous, à des degrés divers, traversés par deux émotions : nous sommes enthousiastes lorsque nous imaginons les progrès qui peuvent découler de l'intelligence artificielle, mais, en même temps, nous sommes effrayés lorsque nous voyons les dangers inhérents à son utilisation [4].

Nous ne pouvons d'ailleurs douter que l'avènement de l'intelligence artificielle représente une véritable révolution cognitivo-industrielle qui contribuera à la création d'un nouveau système social caractérisé par de complexes transformations historiques. Par exemple, l'intelligence artificielle pourrait permettre la démocratisation de l'accès au savoir, le progrès exponentiel de la recherche scientifique, la possibilité de confier des travaux pénibles à des machines ; mais, en même temps, elle pourrait entraîner une plus grande injustice entre les pays riches et les pays en voie de développement, entre les classes sociales dominantes et les classes sociales opprimées, compromettant ainsi la possibilité d'une "culture de la rencontre" au profit d'une "culture du rejet".

L'ampleur de ces transformations complexes est évidemment liée au développement technologique rapide de l'intelligence artificielle elle-même.

C'est précisément cette avancée technologique vigoureuse qui fait de l'intelligence artificielle *un outil fascinant et redoutable* et qui appelle une réflexion à la hauteur de la situation.

Dans ce sens, on pourrait peut-être partir du constat que l'intelligence artificielle est avant tout *un outil*. Et il va de soi que les bienfaits ou les méfaits qu'elle apportera dépendront de son utilisation.

C'est certainement vrai, puisqu'il en a été ainsi pour tous les outils construits par l'homme depuis la nuit des temps.

Notre capacité à construire des outils, en quantité et complexité inégalées parmi les êtres vivants, fait parler d'une *condition techno-humaine* : l'être humain a toujours entretenu une relation avec l'environnement par l'intermédiaire des outils qu'il a progressivement produits. Il n'est pas possible de séparer l'histoire de l'homme et de la civilisation de l'histoire de ces outils. Certains ont voulu lire dans tout cela une sorte de manque, de déficit de l'être humain, comme si, en raison de ce déficit, il était contraint de donner vie à la technique [5]. Un regard attentif et objectif nous montre en fait le contraire. Nous vivons dans une condition d'ultériorité par rapport à notre être biologique ; nous sommes des êtres déséquilibrés par rapport à notre extérieur, voire radicalement ouverts sur l'au-delà. C'est de là que vient notre ouverture aux autres et à Dieu ; c'est de là que naît le potentiel créatif de notre intelligence en termes de culture et de beauté ; c'est de là finalement que provient notre capacité technique. La technologie est donc la trace de cette ultériorité.

Cependant, l'utilisation de nos outils n'est pas toujours uniquement orientée vers le bien. Même si l'être humain sent en lui une vocation à l'au-delà et à la connaissance vécue comme instrument du bien au service des frères et sœurs et de la *maison commune* (cf. *Gaudium et spes*, n. 16), cela ne se produit pas toujours. Au contraire, il n'est pas rare que, précisément à cause de sa liberté radicale, l'humanité ait perverti les finalités de son être en se transformant en son propre ennemi ainsi que de la planète [6]. Les outils technologiques peuvent connaître le même sort. Ce n'est que si leur vocation au service de l'humain est garantie que les outils technologiques révéleront non seulement la grandeur et la dignité unique de l'être humain, mais aussi le mandat qu'il a reçu de "cultiver et garder" (cf. *Gn 2*, 15) la planète et tous ses habitants. Parler de technologie, c'est parler de ce que signifie être humain et de notre condition unique entre liberté et responsabilité, c'est-à-dire parler d'éthique.

Lorsque nos ancêtres aiguisaient des silex pour fabriquer des couteaux, ils les utilisaient à la fois pour couper le cuir pour les vêtements et pour s'entretuer. On pourrait dire la même chose pour d'autres technologies beaucoup plus avancées, comme l'énergie produite par la fusion d'atomes telle qu'elle se produit sur le soleil, qui pourrait certes être utilisée pour produire une énergie propre et renouvelable, mais aussi pour réduire notre planète en un tas de cendres.

L'intelligence artificielle est cependant un outil encore plus complexe. Je dirais presque qu'il s'agit d'un outil *sui generis*. Alors que l'utilisation d'un outil simple (comme le couteau) est sous le contrôle de l'être humain qui l'utilise et que son bon usage ne dépend que de lui, l'intelligence artificielle, en revanche, peut s'adapter de manière autonome à la tâche qui lui est assignée et, si elle est conçue de cette manière, faire des choix indépendants de l'être humain pour atteindre l'objectif fixé [7].

Il faut toujours garder à l'esprit que la machine peut, sous certaines formes et par ces nouveaux moyens, produire des choix algorithmiques. Ce que fait la machine est un choix technique entre plusieurs possibilités et se base soit sur des critères bien définis, soit sur des déductions statistiques. L'être humain, quant à lui, non seulement choisit, mais dans son cœur il est capable de décider. La décision est un élément que nous pourrions concevoir plus stratégique qu'un choix et nécessite une évaluation pratique. Parfois, et bien souvent dans la tâche difficile de gouverner, nous sommes appelés à prendre des décisions qui ont des conséquences pour de nombreuses personnes. Depuis toujours la réflexion humaine parle à ce propos de sagesse, la *phronesis* de la philosophie grecque et au moins en partie la sagesse de l'Écriture Sainte. Face aux prodiges des machines, qui semblent capables de choisir de manière autonome, nous devons être clairs sur le fait que la décision doit toujours être laissée à l'être humain, même dans une tournure dramatique et urgente avec laquelle elle se présente parfois dans nos vies. Nous condamnerions l'humanité à un avenir sans espoir si nous retirions aux gens la capacité de décider d'eux-mêmes et de leur vie, les condamnant à dépendre des choix des machines. Nous devons garantir et protéger un espace de contrôle humain significatif sur le processus de choix des programmes d'intelligence artificielle : la dignité humaine elle-même en dépend.

Permettez-moi d'insister précisément sur ce sujet : dans un drame tel qu'un conflit armé, il est urgent de repenser le développement et l'utilisation de dispositifs tels que les "armes autonomes létales" afin d'en interdire l'usage, en commençant déjà par un engagement dynamique et concret à introduire un contrôle humain de plus en plus significatif. Aucune machine ne devrait jamais choisir d'ôter la vie à un être humain.

Il faut, en outre, ajouter que le bon usage, au moins des formes avancées d'intelligence artificielle, ne sera pas entièrement sous le contrôle des utilisateurs ou des programmeurs qui en ont défini les visées originelles au moment de la conception. Ceci est d'autant plus vrai qu'il est fort probable que, dans un avenir assez proche, les programmes d'intelligence artificielle pourront communiquer directement entre eux afin d'améliorer leurs *performances*. Et si, dans le passé, les hommes qui ont façonné des outils simples ont vu leur existence modifiée par eux – le couteau leur a permis de survivre au froid mais aussi de développer l'art de la guerre –, maintenant que les hommes ont façonné un outil complexe, ils verront celui-ci modifier encore davantage leur existence [8].

Le mécanisme de base de l'intelligence artificielle

Je voudrais maintenant aborder brièvement la complexité de l'intelligence artificielle. Dans son essence, l'intelligence artificielle est un outil conçu pour résoudre un problème et fonctionne par un enchaînement logique d'opérations algébriques, effectuées sur des catégories de données, qui sont confrontées pour découvrir des corrélations, en améliorant leur valeur statistique, grâce à un processus d'auto-apprentissage, basé sur la recherche de nouvelles données et l'auto-modification de ses procédures de calcul.

L'intelligence artificielle est ainsi destinée à résoudre des problèmes spécifiques, mais pour ceux qui l'utilisent, la tentation est souvent irrésistible de tirer, à partir des solutions spécifiques qu'elle propose, des déductions générales, voire anthropologiques.

Un bon exemple est l'utilisation des programmes destinés à aider les magistrats à décider de l'assignation à résidence des détenus purgeant une peine dans un établissement pénitentiaire. Dans ce cas, on demande à l'intelligence artificielle de pronostiquer la probabilité de récidive d'un crime commis par un condamné à partir de catégories prédéfinies (type de crime, comportement en prison, évaluation psychologique et autres), ce qui permet à l'intelligence artificielle d'avoir accès à des catégories de données touchant à la vie privée du condamné (origine ethnique, niveau d'éducation, marge de crédit et autres). L'utilisation d'une telle méthodologie – qui risque parfois de déléguer *de facto* à une machine le dernier mot sur le sort d'une personne – peut renvoyer implicitement aux partialités inhérentes aux catégories de données utilisées par l'intelligence artificielle.

Le fait d'être classé dans un certain groupe ethnique ou, plus prosaïquement, d'avoir commis un délit mineur des années auparavant (ne pas avoir payé, par exemple, une amende pour un

stationnement interdit) influencera, en effet, la décision concernant le fait de procéder ou non à une assignation à résidence. Au contraire, l'être humain évolue en permanence et se montre capable de surprendre par ses actes, chose que la machine ne peut pas prendre en compte.

Il convient également de noter que des applications similaires à celle qui vient d'être mentionnée subiront une accélération du fait que les programmes d'intelligence artificielle seront de plus en plus dotés de la capacité d'interagir directement avec des êtres humains (*chatbots*), en tenant des conversations avec eux et en établissant avec eux des relations étroites, souvent très agréables et rassurantes, car ces programmes d'intelligence artificielle seront conçus pour apprendre à répondre, de manière personnalisée, aux besoins physiques et psychologiques des êtres humains.

Oublier que l'intelligence artificielle n'est pas un autre être humain et qu'elle ne peut proposer de principes généraux, est souvent une grave erreur qui découle du besoin profond de l'être humain de trouver une forme stable de compagnie ou d'un présupposé inconscient de sa part, à savoir que les observations obtenues au moyen d'un mécanisme de calcul sont pourvues des qualités de certitude indiscutable et d'universalité irréfutable.

Cette hypothèse est toutefois risquée, comme le montre l'examen des limites inhérentes au calcul lui-même. L'intelligence artificielle utilise des opérations algébriques à effectuer dans une séquence logique (par exemple, si la valeur de X est supérieure à celle de Y, on multiplie X par Y ; sinon, on divise X par Y). Cette méthode de calcul, appelée "algorithmique", ne présente ni objectivité ni neutralité [9]. En effet, puisque elle est basée sur l'algèbre, elle ne peut examiner que des réalités formulées en termes numériques [10].

Il ne faut pas oublier non plus que les algorithmes conçus pour résoudre des problèmes très complexes sont si sophistiqués qu'il est difficile pour les programmeurs eux-mêmes de comprendre exactement comment ils réussissent à obtenir leurs résultats. Cette tendance à la sophistication risque de s'accélérer considérablement avec l'introduction des ordinateurs quantiques qui ne fonctionnent pas avec des circuits binaires (semi-conducteurs ou puces), mais selon les lois, pour le moins complexes, de la physique quantique. D'autre part, l'introduction continue de puces de plus en plus performantes est déjà devenue l'une des causes de la prépondérance de l'usage de l'intelligence artificielle par les quelques nations qui en sont équipées.

Qu'elles soient sophistiquées ou non, la qualité des réponses fournies par les programmes d'intelligence artificielle dépend en fin de compte des données qu'ils utilisent et de la manière dont elles sont structurées.

Enfin, je voudrais souligner un dernier domaine dans lequel apparaît clairement la complexité du mécanisme de l'intelligence artificielle dite générative (*Generative Artificial Intelligence*). Nul ne

doute qu'il existe aujourd'hui de magnifiques outils d'accès à la connaissance qui permettent même l'auto-apprentissage et l'auto-tutorat dans une myriade de domaines. Beaucoup d'entre nous ont été impressionnés par les applications facilement disponibles en ligne pour composer un texte ou produire une image sur n'importe quel thème ou sujet. Les étudiants se montrent particulièrement attirés par cette perspective qui, lorsqu'ils doivent préparer des travaux, en font un usage disproportionné.

Ces élèves, souvent bien mieux préparés et habitués à l'utilisation de l'intelligence artificielle que leurs professeurs, oublient cependant que l'intelligence artificielle dite générative, au sens strict, n'est pas vraiment "générative". En effet, cette dernière recherche dans les *big data* des informations et les conditionne dans le style qui lui est demandé. Elle ne développe pas de nouveaux concepts ou de nouvelles analyses. Elle répète celles qu'elle trouve, en leur donnant une forme attrayante. Et plus une notion ou une hypothèse est répétée, plus elle la considère comme légitime et valable. Plutôt que "générative", elle est donc "renforçatrice", en ce sens qu'elle réorganise des contenus existants, contribuant à les consolider, souvent sans vérifier s'ils contiennent des erreurs ou des idées préconçues.

Cela risque non seulement de légitimer les *fake news* et de renforcer le poids d'une culture dominante, mais aussi de saper le processus éducatif *in nuce*. L'éducation qui devrait fournir aux étudiants la possibilité d'une réflexion authentique risque de se réduire à une répétition de notions, qui seront de plus en plus estimées incontestables, simplement parce que constamment reprises [11].

Remettre la dignité de la personne au centre d'une proposition éthique partagée

Il convient maintenant d'ajouter une observation plus générale à ce qui a déjà été dit. La saison d'innovation technologique que nous traversons actuellement s'accompagne en effet d'une conjoncture sociale particulière et sans précédent : sur les grandes questions de la vie sociale, il est de plus en plus difficile de trouver des accords. Même dans les communautés caractérisées par une certaine continuité culturelle, des débats passionnés et des confrontations surgissent souvent, rendant difficile la production de réflexions et de solutions politiques partagées visant à rechercher ce qui est bon et juste. Au-delà de la complexité des visions légitimes qui caractérisent la famille humaine, un facteur émerge qui semble unir ces différentes instances. Nous assistons à une disparition ou du moins à une éclipse du sens de l'humain et à une apparente insignifiance du concept de dignité humaine [12]. Il semble que nous perdons la valeur et le sens profond de l'une des catégories fondamentales de l'Occident : la catégorie de la personne humaine. En ce moment où les programmes d'intelligence artificielle remettent en question l'être humain et son agir, c'est précisément la faiblesse de l'*ethos* lié à la perception de la valeur et de la dignité de la personne humaine qui risque d'être le plus grand *vulnus* dans la mise en œuvre et le développement de ces systèmes. En effet, il ne faut pas oublier qu'aucune innovation n'est neutre. La technologie naît dans un but précis et, dans son impact sur la société humaine, elle représente toujours une forme

d'ordre dans les relations sociales et une disposition de pouvoir, permettant à certains d'accomplir des actions et empêchant d'autres d'en accomplir d'autres. Cette dimension constitutive de pouvoir de la technologie comprend toujours, de manière plus ou moins explicite, la vision du monde de ceux qui l'ont conçue et développée.

Cela vaut également pour les programmes d'intelligence artificielle. Pour qu'ils soient des outils pour la construction du bien et d'un avenir meilleur, ils doivent toujours être ordonnés au bien de chaque être humain. Ils doivent avoir une inspiration éthique.

La décision éthique, en effet, est celle qui prend en compte non seulement des résultats d'une action, mais aussi des valeurs en jeu et des devoirs qui en découlent. C'est pourquoi j'ai accueilli favorablement, la signature à Rome, en 2020, du *Rome Call for AI Ethics* [13] et son soutien à cette forme de modération éthique des algorithmes et des programmes d'intelligence artificielle que j'ai appelée "algor-éthique" [14]. Dans un contexte pluriel et global, où s'affichent également des sensibilités différentes et des hiérarchies plurielles dans les échelles de valeurs, il semble difficile de trouver une hiérarchie unique des valeurs. Mais dans l'analyse éthique, nous pouvons également recourir à d'autres types d'outils : si nous peinons à définir un ensemble unique de valeurs globales, nous pouvons toutefois trouver des principes partagés avec lesquels on aborde et résout tout dilemme ou conflit de vie.

C'est la raison pour laquelle est né le *Rome Call* : dans le terme "algor-éthique", est condensée une série de principes qui se révèlent être une plateforme globale et plurielle capable de trouver le soutien des cultures, des religions, des organisations internationales et des grandes entreprises qui sont des acteurs de ce développement.

La politique dont nous avons besoin

Nous ne pouvons donc pas occulter le risque concret, puisqu'il est inhérent à son mécanisme fondamental, que l'intelligence artificielle limite la vision du monde à des réalités exprimables en chiffres et enfermées dans des catégories préconçues, en évinçant l'apport d'autres formes de vérité et en imposant des modèles anthropologiques, socio-économiques et culturels uniformes. Le paradigme technologique incarné par l'intelligence artificielle risque alors de céder la place à un paradigme bien plus dangereux, que j'ai déjà identifié sous le nom de "paradigme technocratique" [15]. Nous ne pouvons pas permettre qu'un outil aussi puissant et indispensable que l'intelligence artificielle renforce un tel paradigme ; au contraire, nous devons faire de l'intelligence artificielle un rempart précisément contre son expansion.

Et c'est précisément là que l'action politique est urgente, comme le rappelle l'encyclique *Fratelli tutti*. Certes, « pour beaucoup de personnes, la politique est aujourd'hui un vilain mot et on ne peut pas ignorer qu'à la base de ce fait, il y a souvent les erreurs, la corruption, l'inefficacité de certains hommes politiques. À cela s'ajoutent les stratégies qui cherchent à affaiblir la politique, à

la remplacer par l'économie ou la soumettre à quelque idéologie. Mais le monde peut-il fonctionner sans la politique ? Peut-il y avoir un chemin approprié vers la fraternité universelle et la paix sociale sans une bonne politique ? » [16].

Notre réponse à ces dernières questions est : non ! La politique est nécessaire ! Je tiens à réaffirmer à cette occasion que « face à tant de formes mesquines de politique et à courte vue [...] la grandeur politique se révèle quand, dans les moments difficiles, on œuvre pour les grands principes et en pensant au bien commun à long terme. Il est très difficile pour le pouvoir politique d'assumer ce devoir dans un projet de Nation et encore davantage dans un projet commun pour l'humanité présente et future » [17].

Mesdames, Messieurs !

Ma réflexion sur les effets de l'intelligence artificielle sur l'avenir de l'humanité nous amène donc à considérer l'importance d'une "saine politique" pour envisager notre avenir avec espoir et confiance. Comme je l'ai déjà dit ailleurs, « sur le plan mondial, la société a de sérieux défauts structurels qu'on ne résout pas avec des rapiècements ou des solutions rapides, purement occasionnelles. Certaines choses sont à changer grâce à des révisions de fond et des transformations importantes. Seule une politique saine sera à même de les conduire, en engageant les secteurs les plus divers et les connaissances les plus variées. De cette manière, une économie intégrée dans un projet politique, social, culturel et populaire visant le bien commun peut "ouvrir le chemin à différentes opportunités qui n'impliquent pas d'arrêter la créativité de l'homme et son rêve de progrès, mais d'orienter cette énergie vers des voies nouvelles" (*Laudato si'*, n. 191) » [18].

C'est précisément le cas de l'intelligence artificielle. Il appartient à chacun d'en faire bon usage et à la politique de créer les conditions pour que cet usage soit possible et fécond.

Merci.

[1] *Message pour la 57ème Journée Mondiale de la Paix du 1er janvier 2024*, n. 1.

[2] Cf. *Ibid.*

[3] Cf. *Ibid.*, n. 2.

[4] Cette ambivalence avait déjà été soulignée par le Pape saint Paul VI dans son *Discours au personnel du "Centre d'automatisation des analyses linguistiques" de l'Aloysianum*, du 19 juin

1964.

[5] Cf. A. Gehlen, *L'uomo. La sua natura e il suo posto nel mondo*, Milano 1983, p. 43.

[6] Cf. Lett. enc. *Laudato si'* (24 mai 2015), nn. 102-114.

[7] *Message pour la 57ème Journée Mondiale de la Paix du 1er janvier 2024*, n. 3.

[8] Les intuitions de Marshall McLuhan et de John M. Culkin sont particulièrement pertinentes en ce qui concerne les conséquences de l'utilisation de l'intelligence artificielle.

[9] Cf. *Discours aux participants à l'Assemblée Plénière de l'Académie Pontificale pour la Vie*, 28 février 2020.

[10] Cf. *Message pour la 57ème Journée Mondiale de la Paix du 1er janvier 2024*, n. 4.

[11] Cf. *Ibid.*, nn. 3 et 7.

[12] Cf. Dicastère pour la Doctrine de la Foi, Déclaration *Dignitas infinita* sur la dignité humaine (2 avril 2024).

[13] Cf. *Discours aux participants à l'Assemblée Plénière de l'Académie Pontificale pour la Vie*, 28 février 2020.

[14] Cf. *Discours aux participants au Congrès "Promoting Digital Child Dignity – From Concept to Action"*, 14 novembre 2019 ; *Discours aux participants à l'Assemblée Plénière de l'Académie Pontificale pour la Vie*, 28 février 2020.

[15] Pour un exposé plus complet, je renvoie à ma Lettre Encyclique *Laudato si'* sur la sauvegarde de la maison commune du 24 mai 2015.

[16] Lett. enc. *Fratelli tutti* sur la fraternité et l'amitié sociale (3 octobre 2020), n. 176.

[17] *Ibid.*, n. 178.

[18] *Ibid.*, n. 179.