



Intelligence artificielle et paix

MESSAGE DE SA SAINTETÉ FRANÇOIS POUR LA 57ème JOURNÉE MONDIALE DE LA PAIX – le 1er JANVIER 2024

En ce début de la nouvelle année, temps de grâce que le Seigneur accorde à chacun d'entre nous, je voudrais m'adresser au Peuple de Dieu, aux nations, aux chefs d'État et de Gouvernement, aux représentants des différentes religions et de la société civile, ainsi qu'à tous les hommes et femmes de notre temps, pour leur présenter mes meilleurs vœux de paix.

1. Le progrès de la science et de la technologie comme chemin vers la paix

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*). L'intelligence est l'expression de la dignité que nous a donnée le Créateur qui nous a créés à son image et à sa ressemblance (cf. *Gn 1, 26*) et nous a permis de répondre à son amour par la liberté et la connaissance. La science et la technologie manifestent de manière particulière cette qualité fondamentalement relationnelle de l'intelligence humaine : elles sont des produits extraordinaires de son potentiel créatif.

Dans la Constitution pastorale *Gaudium et Spes*, le Concile Vatican II a réaffirmé cette vérité en déclarant que « par son travail et son ingéniosité, l'homme a toujours cherché à développer sa propre vie ». [1] Lorsque les êtres humains, « avec l'aide de la technologie », s'efforcent de faire de la terre « une demeure digne de toute la famille humaine », [2] ils agissent selon le plan de Dieu et coopèrent à sa volonté de porter à son achèvement la création et de répandre la paix parmi les peuples. De même, le progrès des sciences et des techniques, dans la mesure où il contribue à un meilleur ordonnancement de la société humaine, à l'accroissement de la liberté et de la communion fraternelle, conduit à l'amélioration de l'homme et à la transformation du monde.

Nous nous réjouissons à juste titre et nous sommes reconnaissants pour les extraordinaires avancées de la science et de la technologie, grâce auxquelles d'innombrables maux qui affligeaient la vie humaine et causaient de grandes souffrances ont été corrigés. En même temps, les progrès techniques et scientifiques, en permettant l'exercice d'un contrôle sans précédent sur la réalité, mettent entre les mains de l'homme un vaste éventail de possibilités, dont certaines peuvent constituer un risque pour la survie de l'humanité et un danger pour la maison commune. [3]

Les remarquables progrès des nouvelles technologies de l'information, en particulier dans la sphère numérique, présentent des opportunités enthousiasmantes et des risques graves, avec de sérieuses implications pour la poursuite de la justice et de l'harmonie entre les peuples. C'est pourquoi il est nécessaire de se poser certaines questions urgentes. Quelles seront les conséquences à moyen et à long terme des nouvelles technologies numériques? Quel sera leur impact sur la vie des individus et de la société, sur la stabilité internationale et sur la paix?

2. L'avenir de l'intelligence artificielle : entre promesses et risques

Les progrès en informatique et le développement des technologies numériques au cours des dernières décennies ont déjà commencé à provoquer de profondes transformations dans la société dans son ensemble, et dans ses dynamiques. Les nouveaux outils numériques changent le visage des communications, de l'administration publique, de l'enseignement, de la consommation, des interactions personnelles et d'innombrables autres aspects de la vie quotidienne.

En outre, les technologies employant une multiplicité d'algorithmes peuvent extraire, à partir des traces numériques laissées sur Internet, des données qui permettent de contrôler les habitudes mentales et relationnelles des personnes, souvent à leur insu, à des fins commerciales ou politiques, en limitant l'exercice conscient de leur liberté de choix. En effet, sur un espace comme la toile, caractérisé par une surcharge d'informations, elles peuvent structurer le flux des données selon des critères de sélection qui ne sont pas toujours perçus par l'utilisateur.

Nous devons rappeler que la recherche scientifique et les innovations technologiques ne sont ni désincarnées de la réalité ni « neutres », [4] mais qu'elles sont soumises à des influences culturelles. En tant qu'activités pleinement humaines, les orientations qu'elles prennent reflètent des choix conditionnés par des valeurs personnelles, sociales et culturelles propres à chaque époque. Il en va de même pour les résultats obtenus : précisément parce qu'ils sont le fruit d'approches spécifiquement

humaines du monde qui les entoure, ils ont toujours une dimension éthique, étroitement liée aux décisions de ceux qui conçoivent l'expérimentation et orientent la production vers des objectifs particuliers.

Il en va de même pour les formes d'intelligence artificielle. Il n'en existe pas à ce jour de définition univoque dans le monde de la science et de la technologie. Le terme lui-même, désormais entré dans le langage courant, englobe une variété de sciences, de théories et de techniques visant à ce que les machines reproduisent ou imitent, dans leur fonctionnement, les capacités cognitives de l'être humain. Parler au pluriel de "formes d'intelligence" permet surtout de souligner le fossé infranchissable qui existe entre ces systèmes, aussi étonnantes et puissantes soient-ils, et la personne humaine : ils sont en définitive "fragmentaires", en ce sens qu'ils ne peuvent qu'imiter ou reproduire certaines fonctions de l'intelligence humaine. L'utilisation du pluriel souligne que ces dispositifs très différents entre eux doivent toujours être considérés comme des "systèmes sociotechniques". En effet, leur impact, quelle que soit la technologie sous-jacente, dépend non seulement de leur conception, mais aussi des objectifs et des intérêts de ceux qui les possèdent et de ceux qui les développent, ainsi que des situations dans lesquelles ils sont utilisés.

L'intelligence artificielle doit donc être comprise comme une galaxie de réalités différentes et nous ne pouvons pas supposer a priori que son développement contribuera de manière bénéfique à l'avenir de l'humanité et à la paix entre les peuples. Un tel résultat positif ne sera possible que si nous nous montrons capables d'agir de manière responsable et de respecter les valeurs humaines fondamentales telles que « l'inclusion, la transparence, la sécurité, l'équité, la confidentialité et la fiabilité ». [5]

Il ne suffit pas non plus de supposer que ceux qui conçoivent les algorithmes et les technologies numériques s'engagent à agir de manière éthique et responsable. Des organismes doivent être renforcés ou, si nécessaire, créés pour examiner les questions éthiques émergentes et protéger les droits de ceux qui utilisent les formes d'intelligence artificielle ou sont influencés par elles. [6]

L'immense expansion de la technologie doit donc s'accompagner d'une formation appropriée à la responsabilité dans son développement. La liberté et la coexistence pacifique sont menacées lorsque les êtres humains succombent à la tentation de l'égoïsme, de l'intérêt personnel, de l'appât du gain et de la soif de pouvoir. Nous avons donc le devoir d'élargir notre regard et d'orienter la recherche technico-scientifique vers la paix et le bien commun, pour le service du développement intégral de l'homme et de la communauté. [7]

La dignité intrinsèque de chaque personne et la fraternité qui nous lient en tant que membres de l'unique famille humaine doivent rester à la base du développement des nouvelles technologies et servir de critères indiscutables pour les évaluer avant leur utilisation, afin que le progrès numérique se fasse dans le respect de la justice et contribue à la cause de la paix. Les développements technologiques qui ne conduisent pas à une amélioration de la qualité de vie de l'ensemble de l'humanité, mais qui au contraire exacerbent les inégalités et les conflits, ne pourront jamais être considérés comme un véritable progrès. [8]

L'intelligence artificielle va devenir de plus en plus importante. Les défis qu'elle pose sont techniques, mais aussi anthropologiques, éducatifs, sociaux et politiques. Elle promet, par exemple, des économies de main-d'œuvre, une production plus efficace, des transports plus faciles et des marchés plus dynamiques, ainsi qu'une révolution dans les processus de collecte, d'organisation et de vérification des données. Nous devons être conscients des transformations rapides en cours et les gérer de manière à sauvegarder les droits humains fondamentaux, en respectant les institutions et les lois qui favorisent le développement humain intégral. L'intelligence artificielle doit servir le potentiel humain le meilleur ainsi que nos aspirations les plus élevées, et non les concurrencer.

3. La technologie du futur : des machines qui apprennent par elles-mêmes

Sous ses diverses formes, l'intelligence artificielle, basée sur des techniques d'apprentissage automatique (*machine learning*), bien qu'elle en soit encore à son stade initial, introduit déjà des changements significatifs dans le tissu des sociétés, exerçant une influence profonde sur les cultures, les comportements sociaux et la construction de la paix.

Des développements tels que l'apprentissage automatique ou l'apprentissage en profondeur (*deep learning*) soulèvent des questions qui dépassent les domaines de la technologie et de l'ingénierie et ont trait à une compréhension étroitement liée au sens de la vie humaine, aux processus fondamentaux de la connaissance et à la capacité de l'esprit à atteindre la vérité.

La capacité de certains appareils à produire des textes syntaxiquement et sémantiquement cohérents, par exemple, n'est pas une garantie de fiabilité. On dit qu'ils peuvent "halluciner", c'est-à-dire générer des affirmations qui semblent à première vue plausibles, mais qui sont en fait infondées ou qui trahissent des préjugés. Cela pose un sérieux problème lorsque l'intelligence artificielle est utilisée dans des campagnes de désinformation qui diffusent des nouvelles fausses et entraînent une méfiance croissante à l'égard des moyens de communication. La confidentialité, la propriété des données et la propriété intellectuelle sont d'autres domaines dans lesquels ces technologies présentent des risques graves, auxquels s'ajoutent d'autres conséquences négatives liées à leur mauvaise utilisation, telles que la discrimination, l'ingérence dans les processus électoraux, la mise en place d'une société qui surveille et contrôle les personnes, l'exclusion numérique et l'exacerbation d'un individualisme de plus en plus déconnecté de la collectivité. Tous ces facteurs risquent d'alimenter les conflits et d'entraver la paix.

4. Le sens de la limite dans le paradigme technocratique

Notre monde est trop vaste, trop diversifié et trop complexe pour être entièrement connu et classifié. L'esprit humain ne pourra jamais en épuiser la richesse, même avec l'aide des algorithmes les plus avancés. Ceux-ci, en effet, ne proposent pas de prévisions garanties de l'avenir, mais seulement des approximations statistiques. Tout ne peut pas être prédit, tout ne peut pas être calculé. En fin de compte, « la réalité est supérieure à l'idée » [9] et, aussi prodigieuse que puisse être notre capacité de calcul, il y aura toujours un résidu inaccessible qui échappera à toute tentative de quantification.

En outre, la grande quantité de données analysées par les intelligences artificielles n'est pas en soi une garantie d'impartialité. Lorsque les algorithmes extrapolent des informations, ils courrent toujours le risque de les déformer, reproduisant les injustices et les préjugés des milieux d'où ils proviennent. Plus ils deviennent rapides et complexes, plus il est difficile de comprendre pourquoi ils ont produit un résultat donné.

Les machines intelligentes peuvent accomplir les tâches qui leur sont assignées avec de plus en plus d'efficacité, mais le but et le sens de leurs opérations continueront à être déterminés ou autorisés par des êtres humains ayant leur propre univers de valeurs. Le risque est que les critères qui sous-tendent certains choix deviennent moins clairs, que la responsabilité de la prise de décision soit dissimulée et que les producteurs puissent se soustraire à l'obligation d'agir pour le bien de la communauté. D'une certaine manière, cela est favorisé par le système technocratique, qui allie l'économie à la technologie et privilie le critère de l'efficacité, tendant à ignorer tout ce qui n'est pas lié à ses intérêts immédiats. [10]

Cela doit nous faire réfléchir sur un aspect très souvent négligé dans la mentalité actuelle, technocratique et recherchant l'efficacité, mais décisif pour le développement personnel et social : le « sens de la limite ». En effet, l'être humain, mortel par définition, pensant dépasser toutes les limites grâce à la technique, risque, dans l'obsession de vouloir tout contrôler, de perdre le contrôle de lui-même ; dans la recherche d'une liberté absolue, de tomber dans la spirale d'une dictature technologique. Reconnaître et accepter ses limites de créature est pour l'homme une condition indispensable pour obtenir, ou mieux accueillir, la plénitude comme un don. Au contraire, dans le contexte idéologique d'un paradigme technocratique, marqué par une présomption prométhéenne d'autosuffisance, les inégalités pourraient croître de manière disproportionnée, le savoir et la richesse s'accumuler dans les mains de quelques-uns, avec de graves risques pour les sociétés démocratiques et la coexistence pacifique. [11]

5. Sujets d'actualité en matière d'éthique

À l'avenir, la fiabilité d'un demandeur de prêt bancaire, l'aptitude d'un individu à un emploi, la possibilité de récidive d'une personne condamnée ou bien le droit à recevoir l'asile politique ou l'aide sociale pourraient être déterminés par des systèmes d'intelligence artificielle. L'absence de divers niveaux de médiation, que ces systèmes présentent, expose particulièrement à des formes de préjugés et de discriminations : les erreurs systémiques peuvent facilement se multiplier, produisant non seulement des injustices dans des cas individuels, mais aussi, par effet domino, de véritables formes d'inégalités sociales.

De plus, les formes d'intelligence artificielle semblent parfois capables d'influencer les décisions des individus par le biais d'options prédéterminées associées à des stimuli et des dissuasions, ou par le biais de systèmes de régulation des choix personnels fondés sur l'organisation des informations. Ces formes de manipulation ou de contrôle social requièrent une attention et une supervision minutieuses et impliquent une responsabilité juridique claire de la part des producteurs, de ceux qui les emploient et des autorités gouvernementales.

Le recours à des processus automatiques qui catégorisent les individus, par exemple par l'utilisation généralisée de la surveillance ou l'adoption de systèmes de crédit social, pourrait également avoir de profondes répercussions sur le tissu de la société, établissant des classements inappropriés entre les citoyens. Ces processus artificiels de classification pourraient également conduire à des conflits de pouvoir, car ils ne concernent pas seulement des destinataires virtuels, mais des personnes en chair et en os. Le respect fondamental de la dignité humaine suppose de refuser que l'unicité de la personne soit identifiée par un ensemble de données. Il ne faut pas permettre aux algorithmes de déterminer la manière dont nous entendons les droits humains, de mettre de côté les valeurs essentielles de compassion, de miséricorde et de pardon, ou d'éliminer la possibilité qu'un individu change et laisse derrière lui le passé.

Dans ce contexte, on ne peut s'empêcher de considérer l'impact des nouvelles technologies dans le domaine du travail : des emplois qui étaient autrefois l'apanage exclusif de la main-d'œuvre humaine sont rapidement absorbés par les applications industrielles de l'intelligence artificielle. Là encore, le risque d'un avantage disproportionné pour quelques-uns au détriment de l'appauvrissement du plus grand nombre est important. Le respect de la dignité des travailleurs et l'importance de l'emploi pour le bien-être économique des personnes, des familles et des sociétés, la sécurité de l'emploi et l'équité des salaires devraient être une priorité absolue pour la Communauté internationale, alors que ces formes de technologies pénètrent de plus en plus profondément sur les lieux de travail.

6. Transformerons-nous les épées en socs ?

En regardant le monde qui nous entoure, on ne peut ces jours-ci échapper aux graves questions éthiques liées au secteur de l'armement. La possibilité de mener des opérations militaires à travers des systèmes de contrôle à distance a conduit à une perception plus faible de la dévastation que ceux-ci causent et de la responsabilité de leur utilisation, contribuant à une

approche encore plus froide et détachée de l'immense tragédie de la guerre. La recherche sur les technologies émergentes dans le domaine des "systèmes d'armes létales autonomes", y compris l'utilisation belliqueuse de l'intelligence artificielle, est un grave sujet de préoccupation éthique. Les systèmes d'armes autonomes ne pourront jamais être des sujets moralement responsables : la capacité humaine exclusive de jugement moral et de décision éthique est plus qu'un ensemble complexe d'algorithmes, et cette capacité ne peut être réduite à la programmation d'une machine qui, bien qu'"intelligente", reste toujours une machine. C'est pourquoi il est impératif de garantir une supervision humaine adéquate, significative et cohérente des systèmes d'armes.

Nous ne pouvons pas non plus ignorer la possibilité que des armes sophistiquées tombent entre de mauvaises mains, facilitant par exemple des attaques terroristes ou des interventions visant à déstabiliser des institutions gouvernementales légitimes. En somme, le monde n'a pas vraiment besoin que les nouvelles technologies contribuent au développement injuste du marché et du commerce des armes, en promouvant la folie de la guerre. Ce faisant, non seulement l'intelligence, mais le cœur même de l'homme, court le risque de devenir de plus en plus "artificiel". Les applications techniques les plus avancées ne doivent pas être utilisées pour faciliter la résolution violente des conflits, mais pour pavé les voies de la paix.

Dans une perspective plus positive, si l'intelligence artificielle était utilisée pour promouvoir le développement humain intégral, elle pourrait introduire d'importantes innovations dans l'agriculture, dans l'éducation et dans la culture, une amélioration du niveau de vie de nations et de peuples entiers, la croissance de la fraternité humaine et de l'amitié sociale. En définitive, la façon dont nous l'utilisons pour inclure les derniers, c'est-à-dire les frères et sœurs les plus faibles et les plus nécessiteux, est la mesure révélatrice de notre humanité.

Un regard humain et le désir d'un avenir meilleur pour notre monde conduisent à la nécessité d'un dialogue interdisciplinaire visant à un développement éthique des algorithmes – l'*'algor-etica'* –, où les valeurs orientent les parcours des nouvelles technologies. [12] Les questions éthiques devraient être prises en compte dès le début de la recherche, ainsi que dans les phases d'expérimentation, de conception, de production, de distribution et de commercialisation. Il s'agit d'une approche de l'éthique de la conception, dans laquelle les institutions éducatives et les décideurs ont un rôle essentiel à jouer.

7. Défis pour l'éducation

Le développement d'une technologie qui respecte et serve la dignité humaine a des implications claires pour les institutions éducatives et pour le monde de la culture. En multipliant les possibilités de communication, les technologies numériques nous ont permis de nous rencontrer de manière nouvelle. Toutefois, une réflexion constante reste nécessaire sur le type de rapports vers lesquels nous nous dirigeons. Les jeunes grandissent dans des environnements culturels imprégnés par la technologie et cela ne peut que remettre en cause les méthodes d'enseignement et de formation.

L'éducation à l'utilisation des formes d'intelligence artificielle devrait viser avant tout à promouvoir la pensée critique. Il est nécessaire que les utilisateurs de tout âge, mais surtout les jeunes, développent une capacité de discernement dans l'utilisation des données et contenus recueillis sur la toile ou produits par des systèmes d'intelligence artificielle. Les écoles, les universités et les sociétés savantes sont appelées à aider les étudiants et les professionnels à s'approprier les aspects sociaux et éthiques du développement et de l'utilisation de la technologie.

La formation à l'utilisation des nouveaux outils de communication devrait tenir compte non seulement de la désinformation, des fausses nouvelles, mais aussi de la recrudescence inquiétante de « peurs ancestrales [...] qui ont su se cacher et se renforcer derrière les nouvelles technologies ». [13] Malheureusement, une fois de plus, nous devons combattre « la tentation de créer une culture de murs, d'élever des murs empêchant la rencontre avec d'autres cultures, avec d'autres personnes » [14] et le développement d'une coexistence pacifique et fraternelle.

8. Défis pour le développement du droit international

Compte tenu de la portée mondiale de l'intelligence artificielle, il est évident qu'à côté de la responsabilité des États souverains de réglementer son utilisation interne, les Organisations internationales peuvent jouer un rôle décisif dans la conclusion d'accords multilatéraux et dans la coordination de leur application et de leur mise en œuvre. [15] À cet égard, j'exhorte la Communauté des nations à travailler ensemble afin d'adopter un traité international contraignant qui réglemente le développement et l'utilisation de l'intelligence artificielle sous ses multiples formes. L'objectif de la réglementation, bien sûr, devrait être non seulement la prévention des mauvaises pratiques, mais aussi l'encouragement des bonnes pratiques, en stimulant des approches nouvelles et créatives et en facilitant des initiatives personnelles et collectives. [16]

En fin de compte, dans la recherche de modèles réglementaires qui puissent fournir un guide éthique aux développeurs de technologies numériques, il est indispensable d'identifier les valeurs humaines qui doivent être à la base de l'engagement des sociétés pour formuler, adopter et mettre en œuvre les cadres législatifs nécessaires. Le travail de rédaction de directives éthiques pour la production de formes d'intelligence artificielle ne peut pas faire abstraction de la prise en compte de questions plus profondes concernant le sens de l'existence humaine, la protection des droits humains fondamentaux, la poursuite de la justice et de la paix. Ce processus de discernement éthique et juridique peut s'avérer être une occasion précieuse pour une réflexion partagée sur le rôle que la technologie devrait avoir dans notre vie individuelle et communautaire, et sur la façon dont son utilisation peut contribuer à la création d'un monde plus équitable et plus humain. C'est pourquoi, dans les débats sur la réglementation de l'intelligence artificielle, il faudrait tenir compte de la voix de toutes les parties prenantes, y compris les

pauvres, les marginalisés et d'autres qui restent souvent ignorés dans les processus décisionnels mondiaux.

J'espère que cette réflexion encouragera à faire en sorte que les progrès dans le développement de formes d'intelligence artificielle servent, en dernière analyse, la cause de la fraternité humaine et de la paix. Ce n'est pas la responsabilité d'un petit nombre, mais de toute la famille humaine. La paix, en effet, est le fruit de relations qui reconnaissent et qui accueillent l'autre dans sa dignité inaliénable, ainsi que de la coopération et de l'engagement dans la recherche du développement intégral de toutes les personnes et de tous les peuples.

Ma prière au début de l'année nouvelle est que le développement rapide de formes d'intelligence artificielle n'augmente pas les trop nombreuses inégalités et injustices déjà présentes dans le monde, mais contribue à mettre fin aux guerres et aux conflits, et à soulager les nombreuses formes de souffrance qui affligen la famille humaine. Puissent les fidèles chrétiens, les croyants de différentes religions et les hommes et les femmes de bonne volonté collaborer en harmonie pour saisir les opportunités et affronter les défis posés par la révolution numérique, et livrer aux générations futures un monde plus solidaire, juste et pacifique.

Du Vatican, le 8 décembre 2023

FRANÇOIS

[1] N. 33.

[2] *Ibid.*, n. 57.

[3] Cf. Lett. enc. *Laudato si'* (24 mai 2015), n. 104.

[4] *Ibid.*, n. 114.

[5] Audience aux participants à la rencontre "Minerva Dialogues" (27 mars 2023).

[6] Cf. *ibid.*

[7] Cf. Message au Président Exécutif du "Forum économique mondial" à Davos-Klosters (12 janvier 2018).

[8] Cf. Lett. enc. *Laudato si'*, n. 194 ; Discours aux participants au Séminaire "Le bien commun à l'ère numérique" (27 septembre 2019).

[9] Exhort. ap. *Evangelii gaudium* (24 novembre 2013), n. 233.

[10] Cf. Lett. enc. *Laudato si'*, n. 54.

[11] Cf. Discours aux participants à l'Assemblée Plénière de l'Académie Pontificale pour la Vie (28 février 2020).

[12] Cf. *ibid.*

[13] Lett. enc. *Fratelli tutti* (3 octobre 2020), n. 27.

[14] *Ibid.*

[15] Cf. *ibid.*, nn. 170-175.

[16] Cf. lett. enc. *Laudato si'*, n. 177.