

Le 3 mai 2025

Lettre ouverte aux cardinaux de l'Église catholique

Objet : Pour une régulation urgente des technologies de l'IA

Éminences,

À l'aube de l'élection du futur pape, nous nous permettons de vous écrire en tant que fidèles catholiques engagés afin d'attirer votre attention sur un sujet d'une importance vitale pour notre société moderne : la régulation des intelligences artificielles (IA).

Il y a bientôt trois ans, l'humanité est entrée dans une nouvelle ère, celle de l'IA générative grand public. Ces nouvelles technologies ne sont pas une avancée de plus. Elles marquent l'entrée dans un nouveau paradigme, celle de la machine autonome. Ces machines apprennent par elles-mêmes à poser des choix complexes que, jusqu'ici, seul l'homme pouvait poser grâce à son intelligence et sa *conscience*.

Antiqua et Nova, publié le 28 janvier dernier, est un document hautement qualitatif qui réaffirme la distinction entre l'homme et la machine face aux frontières que l'IA repousse. Il rappelle très justement que l'homme, contrairement à l'IA, ne doit pas être réduit à un aspect purement fonctionnel. Il sait insister sur le caractère intime propre à l'homme, qui lui permet d'être vrai dans ses actes et ses relations, de vivre librement et pleinement dans toutes les dimensions de son humanité. Sur cet aspect, ce document est un repère profond pour toutes personnes cherchant à mesurer les limites de l'IA.

Cependant, la note soulève une question majeure à laquelle elle ne répond pas : « *à mesure que les modèles d'IA deviennent de plus en plus capables d'apprentissage autonome, la possibilité d'exercer un contrôle sur eux afin de s'assurer que ces applications servent les objectifs humains peut être réduite. Cela pose le problème crucial de savoir comment s'assurer que les systèmes d'IA sont commandés pour le bien des personnes et non contre elles.* » (§45)

La perte progressive du contrôle de ces outils est une réalité que les experts en sécurité de l'IA constatent. Ce constat répond à l'opposé de l'appel réalisé au paragraphe suivant : « *les algorithmes qui pilotent [l'IA] doivent être fiables, sûrs, suffisamment robustes pour gérer les incohérences et transparents dans leur fonctionnement afin d'atténuer les biais et les effets secondaires indésirables* » (§46). Bien que les IA soient entraînées à partir d'algorithmes simples, le produit final de cet entraînement ne constitue pas en soi un algorithme compris ou maîtrisé. Il est pour le moment impossible d'affirmer que ces algorithmes sont « sûrs » et encore moins « transparents dans leur fonctionnement ».

À ce jour, les travaux dans le domaine de l'IA concentrent très peu de ressources pour éclaircir notre compréhension de ces systèmes. Au contraire, des ressources économiques et

humaines considérables sont mises en œuvre pour augmenter leur autonomie. Les IA s'améliorent de jour en jour, les agents autonomes sont en train de se démocratiser, les armes militaires autonomes avancent également très rapidement sans contrôle strict. Enfin, la course à l'intelligence artificielle générale (IAG, une IA largement supérieure aux capacités intellectuelles de l'homme) est activement engagée. Toutes ces avancées tendent à soustraire de ce monde l'homme, favorisant activement notre dépendance totale à cette technologie qui aboutira probablement à en perdre de contrôle.

Si *Antiqua et Nova* rappelle bien que l'*intelligence* de l'IA n'est pas équivalente à l'*intelligence* humaine, cela n'enlève rien au risque soulevé. Le caractère purement réflexif et logique de l'intelligence artificielle n'enlève rien à sa capacité de s'autodévelopper, de se répliquer, d'échapper à notre contrôle et d'agir éventuellement contre l'humanité.

Les experts en sécurité de l'IA sont unanimes : la perte de contrôle de l'IA au détriment de l'humanité dans les années à venir est fortement probable. Les risques qu'une IA échappe à notre contrôle et se retourne contre l'homme sont loin d'être nuls. Pour certains experts, ils sont estimés à environ 10 % dans les deux à trois prochaines décennies. Deux des trois chercheurs au fondement des grands modèles de langages – Geoffrey Hinton et Yoshua Bengio, prix Turing 2018 – font partie de ces lanceurs d'alertes.

L'appel pressant pour réguler les IA n'est à ce jour pas entendu. Les médias ne relaient pas les recherches qui convergent toutes vers ce même résultat. Il semble que la pression des enjeux économiques et du pouvoir associé à l'IA soit trop forte pour qu'un appel à ralentir et à encadrer cette nouvelle industrie soit fructueux.

Ainsi, en ce moment même, une course mondiale à l'IA est engagée entre les nations, qui laissent de côté l'éthique par crainte que l'État adverse ne fabrique une IA supérieure. En effet, un tel outil conférerait au pays qui la détient une puissance stratégique qui mettrait en péril la sécurité des autres nations. Dans ce contexte, il est extrêmement difficile de trouver un terrain d'entente entre les nations. Il semblerait que trouver un accord pour réguler les IA soit impossible. L'humanité paraît livrée à elle-même.

C'est la raison du cri de détresse que nous vous lançons, Éminences, dans cette lettre. En cette période de conclave, nous souhaitons porter votre attention aux enjeux particuliers de l'IA, où l'Église catholique semble avoir un rôle majeur à jouer.

L'Église, par sa foi, sa tradition, son histoire et l'ensemble des principes moraux et philosophiques stables dont elle est dépositaire, lui confère une place et un rôle tout particulier pour la régulation de l'IA à l'échelle internationale. Politiquement, sa neutralité, tant sur le plan diplomatique que sur le plan technique – elle ne développe pas elle-même de modèle de langage – lui permet d'être l'acteur par excellence auprès des autres nations. Enfin, sa défense active de l'homme et du bien commun, inscrite dans son essence même, lui offre une légitimité et une autorité sur ce sujet qu'aucun autre État ne possède.

Pour toutes ces raisons, nous pensons que l'Église catholique devra agir activement, comme elle le fit pendant la guerre froide pour l'arme nucléaire, en faveur d'un accord international pour réguler les IA au profit du bien commun. Il n'est en effet pas exagéré de comparer notre temps à celui de la seconde moitié du siècle dernier, alors que les risques pour l'humanité sont du même ordre de grandeur.

Nous vous prions instamment de porter une attention particulière au contexte actuel pour discerner et choisir celui qui succédera au trône de Saint-Pierre, et qui, faisant face aux défis de l'IA, devra œuvrer, fidèlement à la tradition de l'Église, au bien commun et à la paix sur Terre.

C'est unis par la prière que nous invoquons l'Esprit-Saint pour qu'il vous éclaire et vous guide dans cette tâche. Que votre discernement soit empreint de sagesse et de clairvoyance, afin que l'Église continue d'être un phare de lumière et d'espoir pour l'humanité tout entière.

Nous vous prions d'agréer, Eminences, l'expression de notre profond respect.