



AI

BIG DATA

# DU CŒUR DE L'ÉTAT

## Le Gouvernorat se raconte

Année 2 Cité du Vatican Numéro 3



TRIMESTREIL JUILLET-SEPTEMBRE 2025



Publication du Gouvernorat de l'État de la Cité du Vatican

Communication institutionnelle  
00120 Cité du Vatican  
(État de la Cité du Vatican)

Email: [comunicazione@scv.va](mailto:comunicazione@scv.va)  
Site internet: [www.vaticanstate.va](http://www.vaticanstate.va)

X (Twitter): [Governatorato\\_SCV](#)  
Instagram: [Governatorato\\_SCV](#)

Responsable éditorial: Nicola Gori  
Graphisme et mise en page: Antonio Coretti

Éditeur: Gouvernorat de l'État de la Cité du Vatican



# PAS SEULEMENT LA TECHNOLOGIE

L'Intelligence Artificielle a depuis longtemps cessé d'être de la science-fiction. Aujourd'hui, elle guide des processus, oriente des choix, accompagne des personnes et des organisations dans des contextes toujours plus concrets. Des logiciels qui produisent du contenu, des modèles qui évoluent de manière autonome, des outils capables de suggérer des solutions avant même que le problème ne soit formulé. Un changement puissant, mais pas toujours facile à interpréter.

C'est de cette complexité que naît cette newsletter. Son objectif est d'offrir une boussole dans un écosystème saturé d'annonces, de battage médiatique et de récits simplifiés, en distinguant ce qui compte réellement de ce qui n'est que temporaire. Cette newsletter se veut un lieu de lecture critique et d'approfondissement, destiné à ceux qui souhaitent s'orienter dans l'IA en toute conscience.

Elle accueille des contributions visant à fournir des informations importantes, expliquées dans un langage clair, avec des exemples concrets d'application et des pistes de réflexion sur les conséquences de l'Intelligence Artificielle dans la société, l'économie et la culture. Aucune formation technique spécialisée n'est requise, seulement de l'attention, de la curiosité et l'envie de comprendre.

Parce que l'IA n'est pas seulement une technologie : c'est un prisme à travers lequel nous redéfinissons notre relation avec le futur. Et la manière dont nous en parlons aujourd'hui contribuera à déterminer la façon dont nous la vivrons demain.

Bonne lecture.

**Nicola Gori**  
**La Rédaction**

# L'INTELLIGENCE ARTIFICIELLE : UN OUTIL AU SERVICE DE L'HUMANITÉ

Sœur Raffaella Petrini

Présidente du Gouvernorat de l'État de la Cité du Vatican

Consacrer une édition de la newsletter à l'Intelligence Artificielle est plus que jamais important à une époque où elle se développe et se diffuse toujours davantage.

D'autre part, comme l'a affirmé le Pape François lors de sa participation à la session du G7 sur l'Intelligence Artificielle, le vendredi 14 juin 2024 : « La science et la technologie sont donc des produits extraordinaires du potentiel créatif qui est le nôtre, en tant qu'êtres humains. Or, c'est précisément de l'usage de ce potentiel créatif que Dieu nous a donné que naît l'Intelligence Artificielle ».

En effet, elle est un outil « extrêmement puissant », qui trouve des applications dans de nombreux domaines : de la médecine au monde du travail, de la culture au domaine de la communication, de l'éducation à la politique. Il est évident que l'Intelligence Artificielle aura une influence croissante sur notre vie, sur les relations sociales et sur les rapports entre communautés, institutions et nations.

Face à une réalité aussi complexe et en évolution dynamique, deux attitudes se dessinent : l'enthousiasme pour ses potentialités et la crainte de ses conséquences. Il n'y a aucun doute que l'Intelligence Artificielle doit être gérée et non subie, et elle doit toujours avoir pour objectif la protection de la dignité humaine et servir le bien intégral. De là surgissent une série de questions éthiques concernant son application. Le débat ne peut faire abstraction du fait qu'elle n'est pas un autre être humain.

À ce propos, c'est encore le Pape François qui rappelle aux membres du G7 que : « Les algorithmes conçus pour résoudre des problèmes très complexes sont si sophistiqués qu'il est difficile, même pour leurs propres programmeurs, de comprendre exactement comment ils parviennent à leurs résultats ».

Cela ne doit jamais être oublié, sous peine de réduire notre vision du monde « à des réalités exprimables en chiffres et enfermées dans des catégories préfabriquées », imposant des modèles uniformes, sans la créativité et le discernement propres à l'humanité.

En effet, comme l'a souligné le Pape Léon XIV le vendredi 5 décembre, en s'adressant aux participants à la conférence « *Artificial Intelligence and care of our common home* », la dignité humaine « réside dans la capacité de réfléchir, de choisir librement, d'aimer gratuitement, d'entrer en relation authentique avec autrui ». Dans ce sens, l'Intelligence Artificielle a certes « ouvert de nouveaux horizons pour la créativité, mais elle soulève également des questions préoccupantes quant à ses possibles répercussions sur l'ouverture de l'humanité à la vérité et à la beauté, sur notre capacité à nous émerveiller et à contempler ».

C'est pourquoi cette newsletter peut être l'occasion de réfléchir et d'approfondir la connaissance d'une réalité qui marquera nos vies dans un avenir proche.



# DISCOURS DU PAPE LÉON XIV AUX PARTICIPANTS À UN CONGRÈS SUR « ARTIFICIAL INTELLIGENCE AND CARE OF OUR COMMON HOME »

Organisé par la Fondation Centesimus Annus Pro Pontifice et Strategic Alliance of Catholic Research University,  
Salle du Consistoire, Vendredi 5 décembre 2025



Au nom du Père et du Fils et du Saint Esprit. La paix soit avec vous!

Bonjour, *buongiorno*. C'est un plaisir de vous voir, bienvenue!

Je suis heureux de vous saluer, membres de la Fondation *Centesimus Annus Pro Pontifice* et participants à la *Strategic Alliance of Catholic Research Universities*.

Nous nous rencontrons à l'occasion de la publication de votre recherche sur un thème très important. L'avènement de l'Intelligence Artificielle, en effet, s'accompagne d'un changement rapide et profond de la société, qui touche des caractères essentiels de la personne humaine, tels que la pensée critique, la capacité de discernement, l'apprentissage et la sphère des relations interpersonnelles.

Comment pouvons-nous garantir que le développement de l'Intelligence Artificielle serve véritablement le bien commun et ne soit pas utilisé uniquement pour concentrer la richesse et le pouvoir entre les mains de quelques-uns? Comme vous le savez certainement, le bien le plus précieux sur les marchés aujourd'hui se trouve précisément dans le secteur de l'Intelligence Artificielle. Il s'agit d'une question urgente, dans la mesure où cette technologie a déjà un impact concret sur la vie de millions de personnes, chaque jour et en chaque lieu du monde. Comme nous le rappelle la Doctrine sociale de l'Eglise, et comme cela apparaît clairement du travail interdisciplinaire que vous menez, affronter ce défi exige de se poser une question encore plus radicale: que signifie être humains à notre époque?

L'être humain est appelé à collaborer à l'œuvre de la création, et non pas à être uniquement un consommateur passif de contenus produits par une technologie artificielle. Notre dignité réside dans la capacité à réfléchir, à choisir librement, à aimer gratuitement, à entrer dans une relation authentique avec l'autre. L'Intelligence Artificielle a certainement ouvert de nouveaux horizons pour la créativité, mais elle soulève également des questions préoccupantes en ce qui concerne ses possibles répercussions sur l'ouverture de l'humanité à la vérité et à la beauté, sur notre capacité à nous émerveiller et à contempler. Reconnaître et respecter ce qui caractérise la personne humaine et en garantir la croissance harmonieuse est essentiel pour établir un cadre approprié afin de gérer les implications de l'Intelligence Artificielle.

Et il est important ici de nous arrêter sur une préoccupation qui doit toucher notre cœur: la liberté et la spiritualité de nos enfants et de nos jeunes, avec les possibles conséquences de la technologie sur le développement intellectuel et neurologique. Il faut aider et non entraver les nouvelles générations sur leur chemin vers la maturité et la responsabilité. Le bien-être de la société dépend du fait qu'on leur donne la capacité de développer leurs talents et de répondre aux exigences de l'époque et aux besoins des autres avec un esprit libre et généreux. La possibilité d'accéder à de grandes quantités de données et de connaissances ne doit pas être confondue avec la capacité à en tirer une signification et une valeur. Cette dernière exige également la disponibilité à se confronter avec le mystère et les questions ultimes de notre existence, des réalités souvent marginalisées et même ridiculisées par les modèles culturels et de développement dominants. C'est pourquoi il sera fondamental de permettre aux jeunes d'apprendre à utiliser ces instruments avec leur intelligence personnelle, en s'assurant qu'ils soient ouverts à la recherche de la vérité, à une vie spirituelle et fraternelle, et qu'ils élargissent leurs

rêves et l'horizon de leurs décisions mûres. Nous soutenons leur désir d'être différents et meilleurs, car il n'a jamais été aussi clair qu'aujourd'hui qu'une profonde inversion de route est nécessaire dans notre idée de croissance.

Pour édifier avec nos jeunes un avenir qui, notamment à travers le potentiel de l'Intelligence Artificielle, réalise le bien commun, il est nécessaire de retrouver et de renforcer leur confiance dans la capacité humaine à déterminer l'évolution de ces technologies: une confiance qui est aujourd'hui toujours plus minée par l'idée paralysante que son développement suit un parcours inéluctable. Dans ce but, une action coordonnée et commune est nécessaire, impliquant la politique, les institutions, les entreprises, la finance, l'éducation, la communication, les citoyens et les communautés religieuses. Tous ces acteurs sont appelés à accomplir un engagement commun en assumant cette responsabilité partagée. Un engagement qui précède tout profit et intérêt partisan, toujours plus concentrés entre les mains de quelques-uns. Ce n'est qu'à travers une vaste participation, en donnant la possibilité à toutes les voix, même aux plus humbles, d'être écoutées avec respect qu'il sera possible de réaliser ces objectifs ambitieux. Dans ce contexte, le travail de recherche mené par *Centismus-SACRU* représente une contribution vraiment précieuse.

Chers amis, je vous remercie et je vous encourage à poursuivre avec créativité dans la direction tracée par les Ecritures Saintes et par le Magistère. Que vous accompagnent l'intercession de la Bienheureuse Vierge Marie et la bénédiction apostolique que je vous donne à tous.

---

*L'Osservatore Romano*, Édition Quotidienne, Année CLXV n. 280, vendredi 5 décembre 2025, p. 2.





# INTELLIGENCE ARTIFICIELLE : LE VATICAN TRACE UNE VOIE CONSCIENTE VERS L'AVENIR

Ingénieur Antonino Intersimone

Directeur de la Direction des Télécommunications et des Systèmes d'Information

Nous vivons un moment extraordinaire de l'histoire humaine. L'Intelligence Artificielle, une technologie qui récemment encore semblait relever de la pure imagination, est aujourd'hui pleinement intégrée à notre vie quotidienne. Nous la retrouvons dans nos smartphones, nos ordinateurs personnels et les services que nous utilisons chaque jour. Il s'agit d'une force puissante et, comme toute force puissante, elle exige une gestion attentive, éclairée et responsable. C'est précisément pour cette raison que le Gouvernorat de l'État de la Cité du Vatican a choisi de relever ce défi avec détermination et lucidité.

Au cours des derniers mois, le Gouvernorat a élaboré et adopté un ensemble cohérent de lignes directrices destinées à garantir un usage responsable et éthique de l'Intelligence Artificielle. Ces lignes directrices, officiellement promulguées par le décret n. DCCII du 16 décembre 2024, sont entrées en vigueur le 1<sup>er</sup> janvier 2025, marquant le début d'une nouvelle ère de conscience technologique pour l'État de la Cité du Vatican. Il ne s'agit pas de simples documents formels, mais du résultat concret d'un travail approfondi et partagé, né de la prise de conscience qu'il était devenu nécessaire — et urgent — de définir un cadre clair, précis et faisant autorité pour l'utilisation de cette technologie extraordinaire.

Les lignes directrices vaticanes reposent sur un principe fondamental : l'Intelligence Artificielle possède un potentiel remarquable, capable d'apporter des solutions innovantes à des problèmes complexes et d'améliorer de manière significative la qualité de notre vie. Toutefois, ce potentiel doit être soigneusement équilibré avec le devoir de protéger les droits fondamentaux de chaque personne. Il ne s'agit pas simplement de dire « oui » ou « non » à la technologie, mais de rechercher une voie réfléchie, une approche équilibrée permettant de tirer le meilleur de ce que l'Intelligence Artificielle peut offrir tout en préservant les valeurs les plus importantes.

En d'autres termes, les lignes directrices vaticanes ne se présentent pas comme un frein au développement technologique, mais plutôt comme une « boussole » d'orientation. Cette image de la boussole est particulièrement parlante : de même qu'une boussole aide le voyageur à maintenir le bon cap sans l'empêcher d'avancer, ces lignes directrices entendent orienter l'usage de l'Intelligence Artificielle vers un avenir plus conscient et plus humain, sans

pour autant entraver l'innovation. L'objectif est clair : accroître les capacités humaines grâce à la technologie, sans jamais perdre de vue la valeur centrale de la personne.

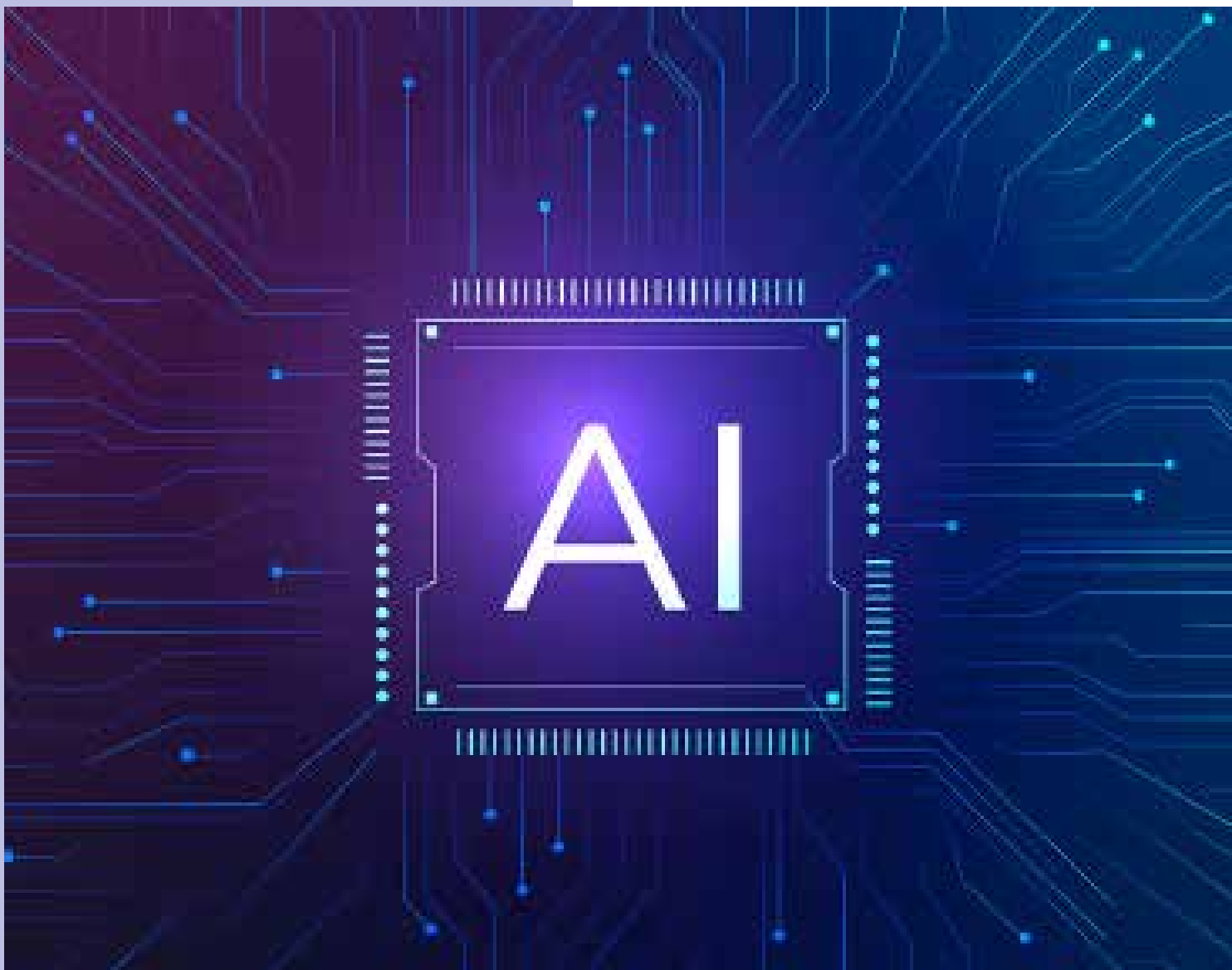


Cette approche reconnaît explicitement le principe de l'irremplaçabilité de l'être humain. Comme l'a rappelé le Pape Léon XIV dans son message adressé aux participants du Congrès international « *Intelligence Artificielle et médecine : le défi de la dignité humaine* » (10–12 novembre 2025), organisé par l'Académie pontificale pour la vie :

*"The objective of providing care for individuals emphasizes the irreplaceable nature of human relationships in this context. Medical professionalism, in fact, requires not only the necessary specific expertise, but also the ability to communicate and be close to others. It can never be reduced merely to solving a problem. Similarly, technological devices must never detract from the personal relationship between patients and healthcare providers. Indeed, if AI is to serve human dignity and the effective provision of healthcare, we must ensure that it truly enhances both interpersonal relationships and the care provided."*<sup>1</sup>

Les lignes directrices abordent un risque bien réel : les systèmes d'Intelligence Artificielle sont susceptibles de perpétuer, voire d'amplifier, les biais déjà présents dans les données sur lesquelles ils sont entraînés. Si, par le passé, les processus de recrutement ont été marqués par des discriminations liées au genre, à l'origine ou à d'autres caractéristiques personnelles, un algorithme formé à partir de ces données historiques pourrait simplement reproduire les mêmes injustices. Le Gouvernement souligne l'importance de prévenir activement ces discriminations potentielles, en veillant à ce que toute procédure de sélection soit non seulement efficace, mais aussi équitable.

C'est précisément pour cette raison que les lignes directrices précisent également que le recours à l'Intelligence Artificielle ne doit ni influencer ni restreindre le pouvoir décisionnel des responsables chargés de l'organisation et de la coordination du personnel. Autrement dit, même lorsque la technologie apporte un soutien, la décision finale revient toujours à une personne — avec sa responsabilité, sa liberté d'appréciation et son discernement. La technologie propose, l'homme dispose.





Les lignes directrices reconnaissent avec honnêteté un fait souvent sous-estimé : malgré ses capacités remarquables, l'Intelligence Artificielle peut causer des préjudices concrets. Il ne s'agit pas de craintes théoriques, mais de risques réels et documentés. Un modèle d'Intelligence Artificielle imparfait peut engendrer des discriminations, porter atteinte aux droits fondamentaux et léser injustement les personnes concernées. C'est pourquoi les lignes directrices accordent une attention particulière à la nécessité de protéger chaque individu contre ces dommages potentiels. Cette protection s'exerce à plusieurs niveaux. D'une part, il y a l'exigence de transparence et d'information : les personnes doivent être en mesure de savoir lorsque l'Intelligence Artificielle influence les décisions qui les concernent. D'autre part, il y a la reconnaissance que certains domaines de la vie humaine — la justice, les droits fondamentaux, la dignité — ne peuvent jamais être entièrement confiés aux algorithmes. Enfin, on trouve l'engagement à suivre, à actualiser et à améliorer continuellement ces lignes directrices au rythme de l'évolution technologique et de l'apparition de nouveaux défis.

Les lignes directrices constituent un document fondamental en raison de l'approche éthique qui les sous-tend. Cette approche reconnaît explicitement un principe qui risque parfois d'être oublié dans l'enthousiasme suscité par le progrès technologique : le principe de l'irremplaçabilité de l'être humain. Aucun algorithme, aussi sophistiqué soit-il, ne peut se substituer pleinement à la sagesse humaine, à l'empathie, ni à la capacité de comprendre le contexte et la signification profonde d'une situation.

Pour cette raison, elles s'imposent comme une référence éthique et normative essentielle pour l'intégration responsable de l'Intelligence Artificielle. Elles représentent une première étape importante vers un avenir dans lequel la technologie est véritablement au service de l'humanité, et non l'inverse. C'est un enseignement qui pourrait — et devrait — être pris en considération avec attention par d'autres institutions, gouvernements et organisations dans le monde entier.

La technologie évoluant à un rythme effréné, les scénarios possibles se multiplient et les défis encore à découvrir sont probablement innombrables. Dans un tel contexte, les lignes directrices ne prétendent pas apporter toutes



les réponses ; elles définissent et établissent les principes fondamentaux. Elles établissent que la justice doit demeurer humaine. Elles établissent que la transparence est un droit et non un privilège. Elles établissent que la dignité humaine n'est pas négociable, même face aux technologies les plus fascinantes.

Le décret du Vatican constitue un modèle significatif pour une réglementation responsable de l'Intelligence Artificielle à l'échelle mondiale. Il n'a pas la prétention d'être un instrument parfait. Il a été élaboré et mis en œuvre avec une attitude attentive, réfléchie et — ce qui est peut-être le plus important — fermement fondée sur la conviction que la technologie doit être au service de l'homme et non l'inverse. Il s'inscrit en parfaite cohérence avec ce qui avait déjà été souligné dans le message adressé à l'*AI for Good Global Summit* du 10 juillet 2025, dans lequel le Saint-Père avait appelé à placer l'Intelligence Artificielle au service de l'humanité tout entière, en rappelant la nécessité de promouvoir la *tranquillitas ordinis*: "*Ultimately, we must never lose sight of the common goal of contributing to that "tranquillitas ordinis – the tranquility of order", as Saint Augustine called it (De Civitate Dei) and fostering a more humane order of social relations, and peaceful and just societies in the service of integral human development and the good of the human family.*"<sup>2</sup>

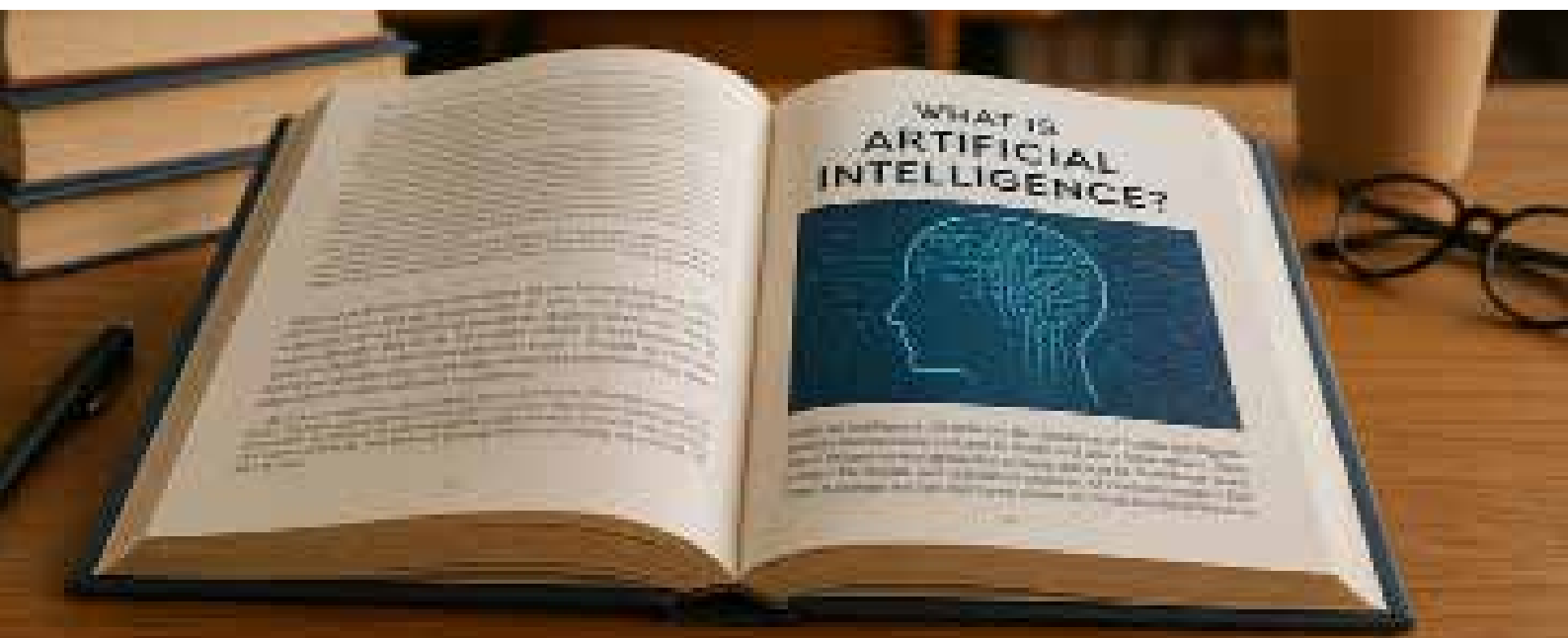
<sup>1</sup> <https://press.vatican.va/content/salastampa/it/bollettino/pubblico/2025/11/10/0854/01523.html>

<sup>2</sup> <https://www.vaticanstate.va/it/novita/2317-messaggio-di-sua-santita-leone-xiv-all-ai-for-good-global-summit-2025.html>

# QU'EST-CE QUE L'INTELLIGENCE ARTIFICIELLE ? UNE EXPLICATION SIMPLE POUR LA COMPRENDRE VRAIMENT

Davide Giordano

Membre de la Commission sur l'Intelligence Artificielle du Gouvenorat



Depuis quelques années, le terme « *Intelligence Artificielle* » a fait une irruption brutale dans la vie de chacun d'entre nous. Nous entendons ce terme à la télévision, dans les journaux télévisés, sur les réseaux sociaux et, bien sûr, même lors de dîners entre amis. Mais qu'est vraiment, pour la plupart d'entre nous, l'Intelligence Artificielle ? D'un point de vue technique et technologique, c'est assurément quelque chose d'extrêmement sophistiqué, impénétrable, réservé à une minorité. Pour tous les autres, l'histoire pourrait être légèrement différente.

Souvent, l'IA est décrite comme « un assistant » qui ne dort jamais, qui ne se fatigue pas, qui connaît des millions et des millions d'informations et qui peut répondre à tes questions en quelques secondes. Cet assistant n'est pas une personne réelle, il n'a pas de corps (ou du moins, pour l'instant, il n'en a pas un fait de tissu organique comme le nôtre), il s'agit uniquement d'un programme situé à l'intérieur d'un ordinateur. Mais il sait beaucoup de choses et parvient à accomplir de nombreuses tâches utiles. Voilà donc le concept de base de ce que nous appelons l'Intelligence Artificielle.

L'Intelligence Artificielle est, en termes très simples, un programme informatique qui a été entraîné à faire des choses intelligentes. Ce n'est pas de la magie, ce n'est pas un mystère : c'est simplement un logiciel, exactement comme les programmes que nous utilisons sur nos téléphones ou nos ordinateurs, sauf que ce logiciel a été entraîné de manière particulière afin de pouvoir comprendre ce que nous lui demandons et de nous fournir des réponses utiles.

Quand nous étions petits, nous avons appris à connaître le monde qui nous entourait par imitation (de maman et papa) ou grâce à des exemples. Pour reconnaître les couleurs, on nous montrait un objet rouge en disant « ceci est rouge », et nous apprenions. Après avoir vu de nombreux objets rouges, notre cerveau comprenait si bien le concept de « rouge » que, lorsque nous voyions un nouvel objet de cette couleur, nous le reconnaissions immédiatement sans que personne n'ait besoin de nous le dire. L'Intelligence Artificielle fonctionne de manière similaire. Les programmeurs apprennent à l'ordinateur à reconnaître certains schémas, certains modèles, en montrant au programme des millions d'exemples. Après avoir vu des millions d'exemples, l'ordinateur apprend le concept et parvient à le reconnaître même lorsqu'on lui présente quelque chose de nouveau.

Voilà le véritable secret de l'Intelligence Artificielle : l'apprentissage, ou *training*. Mais comment un programme apprend-il concrètement ? Ce n'est pas aussi compliqué qu'il n'y paraît. Le processus s'appelle « entraînement », et voici comment il fonctionne de manière simplifiée.

Imaginons que l'on veuille apprendre à un enfant à distinguer un chat d'un chien. On ne lui montre pas un seul chat et un seul chien une seule fois. On lui montre de nombreux chats et de nombreux chiens différents : des chats gris, des chats roux, des chats blancs, des chats grands, des chats petits ; des chiens blancs, des chiens marrons, des chiens petits, des chiens énormes. Après avoir vu des centaines de chats et des centaines de chiens, l'enfant comprend à



quoi ressemble un chat et à quoi ressemble un chien. Lorsqu'il voit un nouvel animal, il est capable de dire s'il s'agit d'un chat ou d'un chien sans que personne ne le lui dise.

L'Intelligence Artificielle fonctionne exactement de la même manière. Les développeurs montrent au programme des millions d'images de chats et des millions d'images de chiens. Ils disent au programme : « Ceci est une image d'un chat » ou bien « Ceci est une image d'un chien ». Le programme observe toutes ces images, repère les modèles communs, identifie les caractéristiques qui font qu'un chat est un chat et qu'un chien est un chien. Après avoir examiné tous ces exemples, si on lui montre une nouvelle photo que le programme n'a jamais vue auparavant, il est malgré tout capable de dire s'il s'agit d'un chat ou d'un chien.

Ce processus d'apprentissage nécessite énormément de données et une très grande puissance de calcul. C'est pourquoi les entreprises qui développent des intelligences artificielles avancées investissent des sommes considérables. Elles ont besoin d'ordinateurs très puissants, qui requièrent de grandes quantités d'énergie pour le traitement et le refroidissement, ainsi que d'énormes volumes de données pour entraîner leurs programmes.

Certains pensent que l'Intelligence Artificielle est quelque chose qui ne sera utilisé que dans des circonstances extraordinaires. En réalité, elle est déjà utilisée plusieurs fois par jour sans même que nous nous en rendions compte.

Lorsque nous écrivons un message sur notre téléphone, celui-ci nous suggère automatiquement le mot suivant que nous pourrions vouloir écrire. C'est de l'Intelligence Artificielle. Le programme a appris, à partir des milliards de messages écrits par les gens, quelles sont les associations de mots les plus courantes. Ainsi, lorsque nous écrivons « Salut, comment », il sait que le mot suivant sera probablement « ça va ».

Lorsque nous activons la reconnaissance faciale sur notre téléphone, le système qui reconnaît notre visage et déverrouille l'appareil est une Intelligence Artificielle. Il a été entraîné sur des millions de visages différents afin d'apprendre à reconnaître les caractéristiques uniques de notre visage.

Lorsque nous utilisons Google ou un autre moteur de recherche, le programme qui décide quels résultats nous afficher est une Intelligence Artificielle. Il a appris, à partir de milliards de recherches précédentes, à comprendre quel résultat correspond le plus probablement à ce que nous cherchons.

Lorsque nous regardons une série télévisée sur Netflix et que le service nous suggère quel épisode regarder ensuite, là encore il s'agit d'Intelligence Artificielle. Lorsque le GPS nous propose l'itinéraire le plus rapide, c'est de l'Intelligence Artificielle. Lorsque l'assistant virtuel, comme Alexa ou Siri, comprend ce que nous disons et nous répond, c'est de l'Intelligence Artificielle.

Nous arrivons maintenant à une distinction importante. Toute l'Intelligence Artificielle n'est pas identique. En réalité, les experts divisent l'Intelligence Artificielle en deux grandes catégories, et comprendre





cette différence nous aidera à mieux saisir le fonctionnement de cette technologie.

Le premier type s'appelle « *Intelligence Artificielle étroite* » (*narrow AI*, ou aussi Intelligence Artificielle faible), et c'est exactement ce que nous avons décrit jusqu'à présent. Il s'agit d'un programme qui est très performant dans l'exécution d'une tâche spécifique. Il excelle dans la reconnaissance faciale, l'analyse de la circulation ou la recommandation de films, mais il ne fait que cela. Si on lui demande d'accomplir quelque chose de différent, il se peut qu'il ne sache pas le faire aussi bien qu'un être humain.

Le second type s'appelle « *Intelligence Artificielle générale* » (AGI, ou aussi Intelligence Artificielle forte ou superintelligence<sup>1</sup>), et c'est quelque chose de très différent. Il s'agirait d'un programme capable de tout faire, exactement comme un être humain. Il pourrait reconnaître des visages, écrire des poèmes, résoudre des problèmes mathématiques complexes, conduire une voiture, cuisiner un plat délicieux... le tout à la perfection.

Ce type d'Intelligence Artificielle n'existe pas encore. Les scientifiques travaillent encore à la théoriser puis à la créer, et beaucoup doutent même que cela soit possible. Pour l'instant, tout ce que nous avons, c'est de l'Intelligence Artificielle étroite, c'est-à-dire des programmes qui sont très performants dans une tâche bien précise.

Lorsque nous lisons dans les journaux qu'il y a une nouvelle découverte en matière d'Intelligence Artificielle, il s'agit presque toujours d'Intelligence Artificielle étroite : un programme qui a été entraîné à bien faire une chose particulière.

Pour bien comprendre ce qu'est l'IA, il est également essentiel de comprendre ce qu'elle ne peut pas faire et quelles sont ses limites.

Premièrement, l'Intelligence Artificielle n'a pas de conscience, ni maintenant ni jamais. Elle ne sait pas qu'elle existe, elle n'a ni sentiments ni désirs. C'est un outil, exactement comme une calculatrice. Une calculatrice est excellente pour faire des calculs mathématiques, mais elle n'a aucune conscience de ce qu'elle fait. Il en va de même pour l'Intelligence Artificielle. Elle ne sait pas qu'elle aide les gens. Elle ne sait pas que son travail est important. Elle se contente d'exécuter les tâches pour lesquelles elle a été programmée.

Deuxièmement, l'Intelligence Artificielle n'est pas créative au sens véritable du terme. Si elle a appris, à partir de millions d'images, que certaines couleurs vont bien ensemble, elle peut combiner ces couleurs d'une manière nouvelle. Mais elle ne crée pas réellement quelque chose de radicalement nouveau : elle se contente de recombinaison des modèles qu'elle a déjà vus. Un artiste humain, en revanche, peut créer quelque chose de totalement inédit, que le monde n'a jamais vu auparavant. Il peut créer à partir de l'inspiration, de l'émotion, d'un élan profond venant de lui-même, ou même d'une erreur : pensez à la manière dont est née la *tarte Tatin*.

Troisièmement, l'Intelligence Artificielle actuelle n'est pas capable de faire un saut logique. Si un nouveau concept est complètement différent de tout ce qu'elle a vu durant son entraînement, elle aura probablement du mal à le comprendre. Un enfant, au contraire, est capable de faire ces sauts logiques et de comprendre des concepts nouveaux en abordant le problème sous des perspectives différentes.

Quatrièmement, l'Intelligence Artificielle peut être facilement trompée. En modifiant légèrement l'image d'un chat de façon à ce que le programme ne le reconnaisse plus comme un chat, alors qu'une personne continuerait à le voir comme tel sans difficulté, l'ordinateur sera déstabilisé. C'est une limite importante : l'Intelligence Artificielle ne « comprend » pas le monde de la même manière que le cerveau humain.

L'Intelligence Artificielle est un outil, ni plus ni moins. Tout comme un marteau, s'il est lancé avec force, il pourra voler mais ne deviendra jamais un avion ; il peut être un instrument utile pour construire un très beau meuble, ou bien être mal utilisé. L'IA est une technologie puissante, capable de résoudre de nombreux problèmes et de rendre notre vie plus confortable et plus efficace. Mais elle n'est pas magique, elle n'est pas consciente, elle n'est ni omnisciente ni omnipotente.

---

<sup>1</sup> Neil Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014)

# INTELLIGENCE ARTIFICIELLE : ENTRE PROMESSES ET DANGERS CONCRETS

Andrea Tripoli

Membre de la Commission sur l'Intelligence Artificielle du Gouvernement

Toute grande innovation technologique porte en elle une double nature : d'un côté, elle offre des possibilités extraordinaires pour améliorer la vie des personnes ; de l'autre, elle introduit des risques et des vulnérabilités qui n'apparaissent souvent que lorsque la technologie est déjà largement diffusée. L'Intelligence Artificielle n'échappe pas à cette règle universelle. Lorsqu'une technologie devient accessible à tous et atteint chaque recoin de la planète, ses potentialités bénéfiques se multiplient, mais les dangers augmentent tout autant. La facilité avec laquelle l'Intelligence Artificielle peut être utilisée en fait un outil extrêmement puissant, aussi bien entre les mains de ceux qui veulent faire le bien que de ceux qui cherchent à l'exploiter à des fins illicites ou nuisibles. La prise de conscience de cette dualité constitue la première étape vers une utilisation responsable et sûre.

L'Intelligence Artificielle fascine profondément de nombreuses personnes. Des programmes qui apprennent

seuls, qui reconnaissent des visages dans une foule, qui rédigent des textes complexes, qui conduisent des voitures sans chauffeur : tout cela semble presque magique. Pourtant, derrière cette apparente perfection technologique se cache une réalité moins rassurante. Les algorithmes d'Intelligence Artificielle peuvent commettre des erreurs importantes, et ces erreurs sont particulièrement dangereuses parce qu'elles ne ressemblent pas à des erreurs. Elles apparaissent comme des décisions rationnelles, objectives, scientifiques, prises par une machine dépourvue d'émotions et de préjugés.

Prenons le cas d'un candidat qui se présente à un entretien d'embauche où le processus de sélection est géré par un système d'Intelligence Artificielle. Le programme évalue les CV, les compétences, la personnalité et le potentiel à l'aide d'algorithmes complexes. En apparence, il s'agit d'un processus équitable et impartial. Pourtant, le système peut commettre des erreurs que personne ne remarque immé-





diatement. Il se peut, par exemple, qu'il ne reconnaisse pas correctement un diplôme parce que la numérisation du document n'est pas parfaite. Ou bien — et c'est encore plus grave — le programme a été entraîné à partir de données historiques de recrutements passés qui reflétaient des discriminations liées au genre, à l'origine ethnique ou à la provenance géographique. Dans ce cas, le système apprend et reproduit ces biais, en excluant automatiquement des candidats parfaitement qualifiés simplement parce qu'ils ne correspondent pas au modèle historique dominant.

Il ne s'agit pas d'un scénario hypothétique. En 2018, Amazon avait développé un système avancé d'Intelligence Artificielle destiné à la sélection de candidats pour des postes techniques. Le programme avait été entraîné en analysant les données des recrutements effectués les années précédentes. Comme, historiquement, les hommes avaient été majoritairement recrutés pour des rôles techniques, l'algorithme a appris que le fait d'être un homme constituait une caractéristique favorable et a commencé à pénaliser systématiquement les candidatures féminines. Le système discriminait automatiquement les femmes, non pas parce qu'il avait été explicitement programmé pour le faire, mais parce qu'il avait appris à partir de données historiques qui reflétaient déjà une discrimination existante. Amazon a été contrainte d'abandonner le projet, mais cet épisode demeure un avertissement particulièrement inquiétant.

Le grand danger réside précisément ici : lorsqu'un ordinateur prend une décision discriminatoire, celle-ci est perçue comme objective et scientifique. Il est difficile de la contester, difficile de démontrer qu'elle est erronée. Si, par le passé, les données utilisées pour entraîner l'Intelligence Artificielle contenaient des discriminations, le programme apprendra ces discriminations et les perpétuera à l'avenir, tout en les revêtant de l'autorité apparente de la science et de la technologie.

Un autre problème fondamental de l'Intelligence Artificielle concerne la transparence, ou plutôt son absence. Il existe, dans le monde technologique, un concept appelé « boîte noire » : on introduit des données en entrée, on obtient des résultats en sortie, mais personne ne sait exactement ce qui se passe entre les deux. Le processus décisionnel qui a conduit à un résultat donné reste inconnu. Lorsque l'Intelligence Artificielle prend des décisions qui influencent profondément la vie des personnes, cette opacité devient un grave problème éthique et pratique.

Supposons qu'une personne demande un prêt bancaire et que la banque utilise un système d'Intelligence Artificielle pour évaluer la demande. Le système analyse des dizaines de variables : revenus, parcours professionnel, zone de résidence, habitudes de consommation, historique des

paiements précédents. À la fin, le système refuse le prêt. Lorsque la personne demande des explications, la banque elle-même est incapable de fournir une réponse précise. Le programme a simplement décidé que le risque était trop élevé, mais même les techniciens qui gèrent le système ne savent pas exactement quels facteurs ont le plus pesé dans la décision, dans quelles proportions, ni selon quelle logique. La décision a été prise, mais elle demeure inexplicable.

Cela crée une injustice intrinsèque profonde. Dans une société démocratique fondée sur la transparence et le droit à la défense, les personnes ont le droit fondamental de savoir pourquoi des décisions qui les concernent directement sont prises. Elles ont le droit de comprendre le raisonnement, de contester d'éventuelles erreurs, de présenter de nouvelles informations susceptibles de modifier l'issue. Or, avec la « boîte noire » de l'Intelligence Artificielle, ce droit est systématiquement bafoué. Les individus doivent accepter des décisions qu'ils ne peuvent pas comprendre, fondées sur des critères qu'ils ne peuvent pas connaître, et contre lesquelles ils disposent de très peu de moyens de défense efficaces.





L'Intelligence Artificielle apprend à partir des données. Cette affirmation simple cache un problème complexe et grave : si les données utilisées pour l'entraînement sont erronées, incomplètes ou reflètent des préjugés existants, l'Intelligence Artificielle apprendra elle aussi des choses fausses et reproduira ces biais à une échelle encore plus large. En 2016, le système COMPAS, utilisé dans les tribunaux américains pour prédire la probabilité qu'un condamné récidive après sa libération, discriminait systématiquement les citoyens afro-américains. Le programme attribuait aux personnes noires un risque de récidive nettement plus élevé qu'aux personnes blanches, même lorsque tous les autres facteurs étaient identiques. La raison était évidente : le système avait été entraîné sur des données historiques reflétant des décennies de discrimination raciale au sein du système de justice pénale américain.

Le problème devient encore plus insidieux lorsque l'on comprend que l'Intelligence Artificielle ne se contente pas de reproduire les préjugés existants : elle les amplifie. Elle fonctionne comme un amplificateur qui prend les défauts de la société et les grossit, les rendant plus omniprésents et plus difficiles à identifier et à corriger. Si la police a historiquement contrôlé plus fréquemment des personnes d'une certaine origine ou d'un certain quartier, un système d'Intelligence Artificielle entraîné sur ces données apprendra que ces personnes sont « plus suspectes » et recommandera de les contrôler encore davantage, créant ainsi un cercle vicieux de discrimination qui s'alimente tout seul.

Alors que, d'un côté, l'Intelligence Artificielle peut commettre des erreurs discriminatoires de manière involontaire, de l'autre, elle peut être délibérément utilisée comme un outil de tromperie et de manipulation. Le paysage de la cybersécurité a été profondément transformé par cette technologie, offrant aux cybercriminels des outils extrêmement puissants pour orchestrer des attaques toujours plus sophistiquées.

Les escroqueries par courrier électronique, connues sous le nom de *phishing*, ont connu une évolution spectaculaire. Les messages maladroits, truffés de fautes de grammaire et d'orthographe qui caractérisaient les arnaques du passé, sont désormais obsolètes. L'Intelligence Artificielle générative permet de créer des communications parfaites dans n'importe quelle langue, totalement indiscernables des messages authentiques. Les escrocs peuvent générer des courriels qui semblent provenir de banques, d'employeurs, de services gouvernementaux ou de plateformes utilisées au quotidien. Le niveau de personnalisation est alarmant : les algorithmes analysent les profils sur les réseaux sociaux, les habitudes d'achat en ligne, les centres d'intérêt personnels et construisent des messages sur mesure qui exploitent les vulnérabilités spécifiques de chaque individu.

Parallèlement, le *smishing*, c'est-à-dire le *phishing* par SMS, est devenu tout aussi dangereux. Des messages avertissant de la livraison d'un colis, de problèmes urgents liés au compte bancaire, de bonus à récupérer ou d'amendes à payer arrivent aux moments les plus opportuns, rédigés de manière impeccable, en exploitant le sentiment d'urgence et la confiance instinctive que beaucoup de personnes accordent aux communications téléphoniques. La brièveté du format SMS, qui rendait autrefois plus difficile la création d'arnaques convaincantes, n'est désormais plus un obstacle pour l'Intelligence Artificielle.

Mais les menaces vont bien au-delà des messages écrits. Les *deepfakes* audio et vidéo représentent une frontière encore plus inquiétante et dangereuse. Avec seulement quelques secondes d'enregistrement vocal, l'Intelligence Artificielle peut cloner parfaitement la voix de n'importe quelle personne. Un appel provenant prétendument de son supérieur demandant d'effectuer en urgence un virement non autorisé, ou d'un membre de la famille en difficulté suppliant une aide financière immédiate : distinguer le vrai du faux devient pratiquement impossible sans vérification indépendante. Des cas d'escroqueries portant sur des millions d'euros orchestrées grâce à cette technologie ont déjà été documentés, et le phénomène est en rapide expansion.

Même les escroqueries téléphoniques traditionnelles, connues sous le nom de *vishing*, ont connu une évolution spectaculaire. Des *chatbots* dotés d'Intelligence Artificielle peuvent mener des conversations naturelles, répondre aux questions en temps réel, manifester une empathie apparente et instaurer un climat de confiance, le tout sans aucune intervention humaine. Ces systèmes fonctionnent vingt-quatre heures sur vingt-quatre, peuvent appeler des milliers de per-



sonnes simultanément et adaptent dynamiquement leur approche en fonction des réponses reçues, devenant de plus en plus convaincants au fil de la conversation.

Face à ce panorama complexe de risques et de menaces, il est essentiel de préciser que le problème n'est pas l'Intelligence Artificielle en elle-même, mais la manière dont elle est développée, diffusée et utilisée. La technologie en soi est neutre ; ce sont les choix humains qui la rendent bénéfique ou nuisible. La bonne nouvelle est que l'Intelligence Artificielle est également utilisée pour protéger les personnes : les systèmes de sécurité modernes analysent des milliards de données afin d'identifier des schémas suspects, de bloquer des cyberattaques en temps réel et de prévoir de nouvelles menaces avant qu'elles ne se concrétisent. Une véritable course aux armements technologique est en cours entre ceux qui utilisent l'Intelligence Artificielle pour attaquer et ceux qui l'utilisent pour défendre.

La prise de conscience constitue la première ligne de défense, la plus importante aussi. Comprendre que l'Intelligence Artificielle peut être utilisée pour créer des escroqueries parfaitement crédibles permet de développer un scepticisme sain à l'égard de communications apparem-

ment officielles. Savoir que les algorithmes peuvent discriminer implique d'exiger de la transparence dans les décisions automatisées qui nous concernent. Reconnaître que les systèmes peuvent être entraînés sur des données biaisées signifie réclamer des contrôles rigoureux avant leur déploiement dans des secteurs critiques tels que la justice, la santé ou la finance.

Vérifier systématiquement l'identité de toute personne qui nous contacte par des canaux indépendants, se méfier des demandes urgentes exigeant une action immédiate, contrôler attentivement les adresses électroniques et les liens avant toute interaction, activer des systèmes d'authentification renforcée : ces pratiques deviennent indispensables dans un monde où l'Intelligence Artificielle peut imiter à la perfection n'importe quelle personne ou institution. L'esprit critique individuel, associé à une réglementation forte et à de solides mécanismes de contrôle, constitue la seule véritable défense efficace contre les risques liés à l'Intelligence Artificielle. En définitive, la technologie doit rester un outil au service de l'humanité, contrôlé par l'humanité, et ne pas se transformer en une force incontrôlable décidant de manière autonome du destin des individus.



# LES LARGE LANGUAGE MODELS : DU CHOIX DES MODÈLES EXISTANTS À L'INVESTISSEMENT DANS DES MODÈLES PERSONNALISÉS

Domenico Vetere

Vice-responsable du Bureau Interne

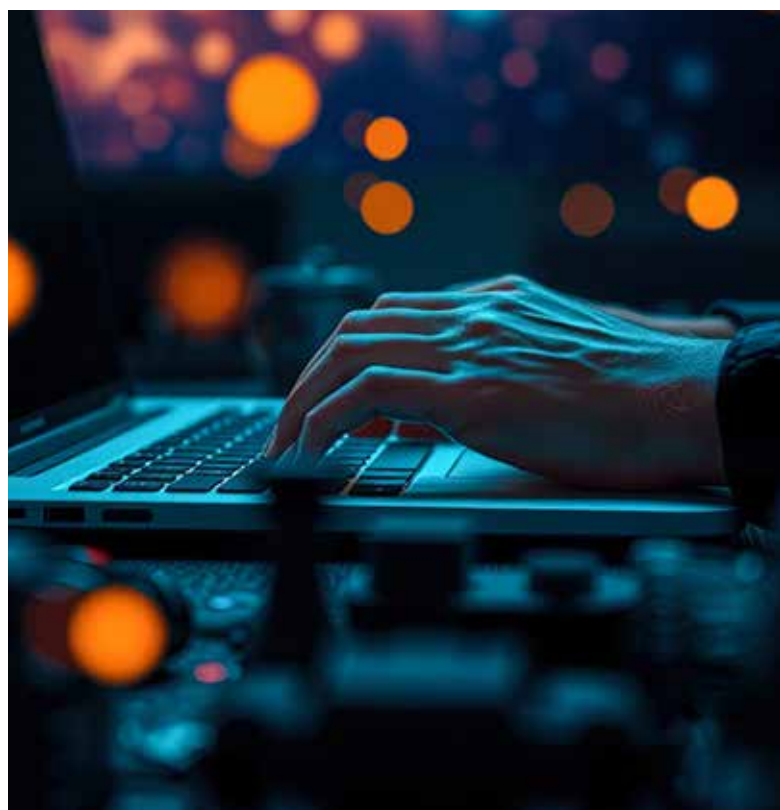
## LARGE LANGUAGE MODEL AI

Lorsque l'on parle d'Intelligence Artificielle dans le contexte organisationnel contemporain, on est de plus en plus souvent confronté aux soi-disant *Large Language Models* (plus connus comme LLM). Ces systèmes ont commencé à transformer notre façon de travailler, de communiquer et de traiter l'information. Cependant, la compréhension de ce qu'ils sont, des raisons pour lesquelles il en existe différents types, de la manière de les choisir et, surtout, des raisons d'investir dans des modèles personnalisés reste encore assez floue dans la majorité des organisations. Cette réflexion découle de la nécessité de clarifier un aspect crucial : tous les LLM ne se valent pas et le choix d'utiliser un modèle existant plutôt que d'investir dans la création d'un modèle propriétaire constitue une décision stratégique fondamentale pour toute institution moderne. Un *Large Language Model* est un système d'Intelligence Artificielle entraîné sur d'énormes quantités de texte (*datasets*), capable de comprendre le langage naturel et de générer des réponses cohérentes, contextualisées et, idéalement, précises. Leur « largeur » — le terme « *large* » — fait référence à la fois au volume de données sur lequel ils sont entraînés et à la complexité des modèles eux-mêmes.

Ces dernières années, le marché s'est ainsi peuplé de LLM très différents les uns des autres. Ce n'est pas le fruit du hasard. Plusieurs organisations — OpenAI avec ChatGPT, Google avec Gemini, Meta avec Llama, Anthropic avec Claude — ont investi massivement dans la création de modèles génériques, car ils représentent une opportunité économique et stratégique significative. Chacun de ces acteurs a fait des choix différents : certains ont opté pour un accès ouvert, d'autres conservent leurs modèles comme propriété exclusive et d'autres encore proposent des versions gratuites et des versions payantes. La raison de cette prolifération est

simple : il n'existe pas de LLM universellement optimal pour tous les contextes. Un modèle qui fonctionne parfaitement pour des traductions littéraires pourrait ne pas être idéal pour l'analyse de données de santé. Un système robuste pour l'écriture créative pourrait ne pas être fiable pour la gestion d'informations confidentielles. Et un modèle entraîné sur des données publiques, aussi sophistiqué soit-il, pourrait ne pas comprendre correctement les langages spécialisés, les protocoles internes ou les nuances culturelles propres à une organisation spécifique.

Si l'on décide d'utiliser un LLM « *off-the-shelf* » (c'est-à-dire un modèle déjà existant et disponible sur le marché), nous faisons face à une variété d'options, chacune présentant des caractéristiques distinctives.





**GPT-5 d'OpenAI**, connu de tous sous le nom de **ChatGPT**, est probablement le plus célèbre. Il est extrêmement polyvalent, capable de traiter une très large gamme de tâches, de la génération de texte et d'images à la résolution de problèmes complexes. Son inconvénient est d'être propriétaire, de nécessiter une connexion internet permanente et de ne pas garantir la confidentialité totale des données saisies (utilisées pour améliorer le modèle lui-même).

**Gemini de Google** offre des fonctionnalités similaires, avec l'avantage supplémentaire d'une intégration étroite aux écosystèmes Google et d'une bonne capacité de traitement multimodal (texte, images, vidéos). Là encore, nous restons dépendants de l'infrastructure cloud de Google et de ses politiques de confidentialité.

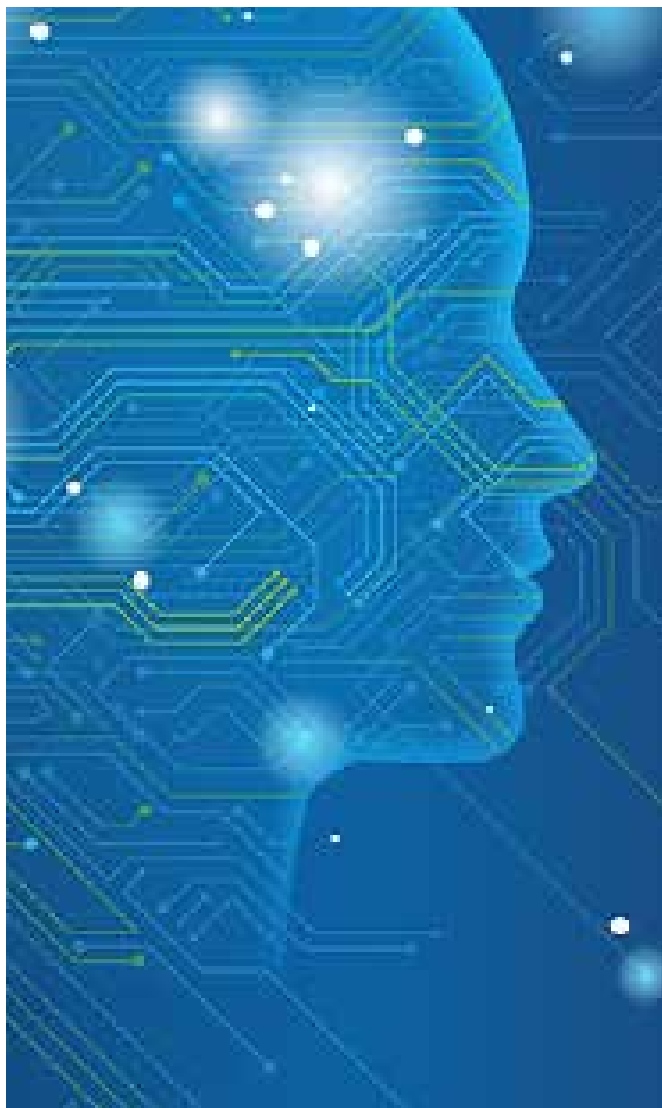
**Claude d'Anthropic** a été conçu en insistant sur la sécurité et la fiabilité, particulièrement dans la gestion d'informations sensibles. Il a été entraîné avec des méthodologies mettant l'accent sur la réduction des risques et la transparence. Il reste néanmoins un modèle externe, avec les limitations que cela implique.

**Llama de Meta** représente une alternative intéressante, car il est disponible en version *open-source*, ce qui signifie que les organisations peuvent le déployer dans des environnements *self-hosted* (c'est-à-dire sur leurs propres serveurs), en gardant un contrôle total sur les données, en respectant pleinement le concept de « souveraineté des données ». Cependant, sa mise en œuvre correcte nécessite des compétences techniques importantes.

Comment choisir entre ces modèles ? Les critères devraient inclure, par exemple : le type de tâches que nous souhaitons confier au modèle, le niveau de sensibilité des données qu'il traitera, la nécessité d'opérer hors ligne ou avec des données confidentielles, le budget disponible, les compétences techniques internes, et le degré de contrôle et de personnalisation que nous souhaitons. Dans cette optique, pour de nombreuses organisations à une phase initiale, l'utilisation d'un LLM public déjà établi représente une solution pragmatique : il est rapide à mettre en œuvre, relativement économique et ne nécessite pas d'investissements importants dans des infrastructures propres. Cependant, arrivé à ce stade de la réflexion, il faut affronter une vérité dérangeante : les LLM génériques, aussi sophistiqués soient-ils, présentent des limites significatives lorsqu'il s'agit d'opérer dans des contextes spécialisés, confidentiels ou hautement personnalisés. Par exemple, imaginez vouloir utiliser un LLM pour assister un bureau qui gère des procédures administratives complexes, basées sur des réglementations spécifiques, des précédents institutionnels ou des interprétations uniques à l'organisation. Un LLM générique aura été entraîné sur une vaste variété de textes publics, mais n'aura jamais vu



les documents internes de l'organisation ; il ne connaît pas les procédures et ne comprend pas les nuances de la culture organisationnelle spécifique. Le risque est double. Premièrement, le modèle pourrait fournir, avec une forte probabilité, des réponses plausibles mais incorrectes ; partager des informations qu'il ne connaît pas, ou interpréter des ambiguïtés d'une manière qui ne reflète pas l'intention de l'organisation. Deuxièmement, pour obtenir des *performances* acceptables en termes de précision, on pourrait être contraint d'intégrer des documents confidentiels, des procédures internes et des informations propriétaires dans des modèles publics, mettant ainsi en danger le capital intellectuel. Ce deuxième point est particulièrement critique. Si une organisation intègre ses données propriétaires dans un LLM public pour obtenir de meilleurs résultats, elle cède en réalité le contrôle de ces informations. Même si des clauses de confidentialité existent, les données contribueront à l'entraînement du modèle, pourront probablement être utilisées par d'autres, et deviendront partie de l'infrastructure technologique d'un tiers. Pour une institution comme le Gouvernorat, ou pour toute organisation ayant des responsabilités vis-à-vis d'informations sensibles, cela représente un risque inacceptable.



C'est pourquoi, pour de nombreuses organisations, l'investissement dans des *Large Language Models* personnalisés — des modèles créés spécifiquement, entraînés exclusivement sur des données propriétaires et totalement contrôlés — représente une option stratégique qui mérite une considération sérieuse. Ainsi, un *LLM personnalisé* est un modèle qui est « *fine-tuned* » (c'est-à-dire ajusté et spécialisé), en partant d'un modèle de base déjà existant et en l'entraînant ensuite sur des données spécifiques. Ce processus présente plusieurs avantages significatifs. Tout d'abord, la propriété intellectuelle continue à appartenir à l'organisation. Les données utilisées pour entraîner le modèle ne sont pas partagées, ne contribuent pas à des modèles publics et restent **protégées**. Cela est crucial pour des informations représentant une valeur concurrentielle, des connaissances historiques uniques ou des données sensibles. Deuxièmement, la *performance* est incroyablement meilleure. Un modèle entraîné sur des documents, des terminologies, des procédures et des styles de communication spécifiques fonctionnera beaucoup mieux dans un contexte interne spécialisé qu'un modèle générique. Par exemple, si un service gère une législation complexe, un modèle entraîné sur des décennies de précédents et d'interprétations propres à l'organisation sera incomparablement plus fiable. Troisième point important, il sera possible de décider exactement sur quelles données entraîner le modèle, quelles sources exclure, comment interpréter les ambiguïtés, et quelles sorties considérer comme acceptables. Tout reste à l'intérieur de l'organisation. Le quatrième aspect à considérer est la sécurité, qui devient implicitement garantie. Un *LLM personnalisé* peut être hébergé sur des serveurs locaux, dans des environnements hors ligne si nécessaire, avec des niveaux de chiffrement et de contrôle entièrement personnalisables. Il n'y a aucune transmission de données vers des infrastructures externes. Le cinquième et dernier point de cette réflexion concerne la durabilité assurée. Un modèle public peut changer, disparaître, être acquis ou modifier ses politiques. Un modèle personnalisé et privé reste la propriété de l'organisation, pour toujours, et peut être adapté selon les besoins futurs.

Cependant, la valeur d'un *LLM personnalisé* dépend inévitablement de la qualité des données utilisées pour son entraînement. C'est ici qu'intervient un aspect souvent sous-estimé : la nécessité d'investir dans la préparation et la certification des données. Comme nous l'avons déjà dit, tous les modèles ne se valent pas. Si l'on entraîne un LLM sur un ensemble désordonné, non vérifié, rempli de contradictions et d'imprécisions, on obtiendra un modèle qui reproduit toutes ces problématiques. L'IA amplifie les biais présents dans les données d'entraînement. Si les documents contiennent des erreurs, le modèle apprendra à les reproduire. Si les informations sont conflictuelles, le modèle deviendra confus et peu fiable. Pour cette rai-



son, un véritable investissement dans un *LLM personnalisé* nécessite également un investissement parallèle dans la gestion et la certification des données : les « *données propres* » ou *données de qualité*.

Cela signifie identifier et sélectionner les données les plus importantes et **de qualité**. Tous les documents historiques ne sont pas nécessairement appropriés. Certains peuvent être obsolètes, d'autres contenir des informations incorrectes, et d'autres encore être des doublons ou redondants. La vérification est essentielle. Il faut contrôler l'exactitude et la cohérence. Avant d'utiliser les données pour l'entraînement, elles doivent être soumises à un contrôle qualité. Cela peut impliquer d'avoir des experts qui vérifient manuellement des échantillons de documents, qui résolvent les contradictions et corrigent les erreurs manifestes. Concernant les sources et les versions, il doit apparaître clairement d'où proviennent les données, quelle version représente la « vérité » officielle, et comment elles ont été transformées. Cette traçabilité est fondamentale pour la fiabilité du modèle. Ensuite, il convient de se poser une autre série de questions : qui a l'autorité de décider quelles données inclure ? Comment les mises à jour sont-elles gérées ? Quel est le cycle de vie d'une donnée ? Ces questions doivent avoir des réponses claires et documentées.

À ce stade, créer un *LLM personnalisé* vaut-il vraiment l'investissement ? La réponse dépend de la situation spécifique, mais pour de nombreuses organisations de taille moyenne ou grande, la réponse est un oui catégorique.

L'effort initial pour développer un *LLM personnalisé* varie selon la complexité du domaine, le volume de données

et le niveau de personnalisation souhaité. Cependant, cet effort doit être considéré dans une perspective à long terme. Si l'on utilise un LLM public en y intégrant continuellement des données confidentielles pour obtenir des performances acceptables, on paie déjà un prix caché : la perte de contrôle sur sa propriété intellectuelle. Si un modèle générique fournit des réponses inexactes et qu'il faut mobiliser des ressources humaines pour corriger continuellement les erreurs, on paie par l'inefficacité. Dépendre d'un fournisseur externe pour chaque opération rend également vulnérable aux changements de tarification, aux interruptions de service et aux modifications unilatérales des politiques. Un *LLM personnalisé*, en revanche, représente *un asset dont la valeur croît dans le temps*. Au fur et à mesure de son utilisation, il peut être constamment amélioré avec de nouvelles données. Il devient progressivement plus intelligent dans votre contexte spécifique. Et il reste, en toute circonstance, votre propriété.

Comme nous l'avons déjà dit, pour de nombreuses organisations, surtout à une phase initiale, l'utilisation pragmatique de modèles publics bien choisis représente une option raisonnable. Cependant, dès lors qu'une organisation constate que l'IA est réellement utile pour ses opérations, qu'elle l'utilise régulièrement et que la qualité des résultats est critique pour son fonctionnement, la donne change. À ce stade, investir dans un *LLM personnalisé* — entraîné sur les données certifiées, propriétaires et propres de l'organisation — n'est pas un luxe, mais une nécessité stratégique.



# L'INTELLIGENCE ARTIFICIELLE AU GOUVERNORAT : LES POLITIQUES DE FORMATION

La croissance exponentielle du phénomène de l'« *Intelligence Artificielle* » est à la fois irréversible et indéniable. L'Intelligence Artificielle constitue une véritable révolution, un bouleversement profond qui pénètre simultanément dans les bureaux, les organisations, les établissements scolaires et les familles. C'est précisément pour cette raison que le Governorat de l'État de la Cité du Vatican a élaboré une stratégie réfléchie et responsable pour relever ce défi historique.

Lorsque l'Intelligence Artificielle entre dans une société, elle ne le fait pas de manière uniforme. Chaque contexte — professionnel, éducatif ou familial — vit cette transformation de façon distincte, avec des défis spécifiques et des opportunités particulières. Comprendre cette diversité est essentiel pour élaborer des réponses appropriées et conscientes.

Dans le contexte professionnel, l'introduction de l'Intelligence Artificielle requiert une vigilance sur de multiples fronts. Il est essentiel de contrôler rigoureusement l'authenticité des contenus générés par l'IA, tout en protégeant la créativité humaine, qui demeure irremplaçable. Il convient également de faire face à des risques émergents tels que la « *shadow AI* », c'est-à-dire l'utilisation de systèmes d'Intelligence Artificielle non autorisés ou non contrôlés opérant dans les interstices des processus de travail. L'adoption de vérifications attentives des processus d'entreprise et de la protection des données, qui représentent aujourd'hui des assets de plus en plus critiques et précieux pour toute organisation moderne, devient indispensable.

Dans le contexte scolaire et éducatif, le défi est encore plus profond. Il est nécessaire de repenser complètement le modèle pédagogique traditionnel : depuis l'offre de formation jusqu'aux méthodes d'enseignement qui ont caractérisé l'éducation jusqu'à aujourd'hui. La communauté éducative s'interroge légitimement sur la manière dont l'Intelligence Artificielle influence la capacité critique des élèves. Certains

experts craignent que la possibilité de déléguer des activités de recherche et de traitement à l'IA puisse appauvrir la pensée critique des jeunes, essentielle au développement de citoyens conscients et autonomes.

Dans le contexte le plus intime et délicat — celui de la famille — l'Intelligence Artificielle touche des aspects beaucoup plus vulnérables de la société. Parents et adultes sont amenés à utiliser des *chatbots* et des assistants virtuels pour traiter des questions qui relevaient traditionnellement des psychologues, des professionnels de l'éducation et des spécialistes. Cela se produit à un moment où l'âge moyen de la première utilisation d'un *smartphone* par un enfant est d'environ dix ans. La question cruciale devient : comment protéger les mineurs lorsqu'ils sont exposés à des technologies qu'ils ne comprennent pas pleinement, et comment guider les adultes dans l'usage responsable de ces outils dans le contexte familial ?

Face à ces interrogations multiples, une réponse commune apparaît clairement : il est nécessaire de créer une conscience spécifique à travers la sensibilisation et l'éducation

numérique. En d'autres termes, la formation devient l'outil principal pour transformer la relation entre les personnes et la technologie.

Le Governorat a reconnu cette nécessité et, depuis plus d'un an, investit de manière significative dans des programmes

de formation spécifiques sur l'Intelligence Artificielle. Ces programmes ne se limitent pas à expliquer comment utiliser les outils, mais abordent des thèmes complexes et structurés. Ils incluent l'histoire du développement de l'Intelligence Artificielle, avec ses cycles d'optimisme et de scepticisme — les fameux « hivers et printemps de l'IA ». Ils traitent des limites intrinsèques de la technologie, des attentes réalistes, des potentiels réels et des erreurs auxquelles l'Intelligence Artificielle est naturellement exposée.



Le catalogue de formation est structuré sur plusieurs niveaux, chacun conçu pour répondre à des besoins différents. Le niveau de base s'adresse à ceux qui font leurs premiers pas dans le monde de l'Intelligence Artificielle et souhaitent se familiariser avec la terminologie et les contextes d'application. Les niveaux plus avancés permettent un contact direct avec les différentes plateformes et les différents modèles de langage (LLM). Grâce à une expérience pratique, les participants commencent à comprendre comment identifier l'outil le plus approprié en fonction de l'objectif spécifique qu'ils souhaitent atteindre.

De par la nature même du sujet, les programmes et contenus de formation ne sont pas statiques. Compte tenu du rythme accéléré du changement technologique, le catalogue de formation est mis à jour de manière constante, reflétant toujours la dynamique propre au domaine. Cette approche flexible garantit que la formation reste pertinente et actuelle.

Le Gouvernorat ne relève pas seul le défi de l'IA. L'Intelligence Artificielle est un phénomène mondial qui nécessite des solutions coordonnées à l'échelle internationale. C'est pourquoi le Gouvernorat est en lien actif avec des organisations internationales capables de réunir des centaines de pays pour dialoguer sur des questions technologiques critiques.

Parmi ces initiatives, la participation au *World Summit on the Information Society* (WSIS) et à ses prolongements récents revêt une importance particulière. Dans ce cadre, le Gouvernorat est impliqué dans des projets tels que « *AI for Good* », un événement et une plateforme visant à identifier des technologies, méthodes et stratégies visant à l'adoption consciente et responsable de l'Intelligence Artificielle dans tous les contextes d'application.

Ces stratégies globales abordent nécessairement des problèmes déjà connus et bien documentés. Le premier d'entre eux est le fossé numérique (*digital divide*), c'est-à-dire l'accès inégal à la technologie entre les pays riches et les pays en développement. En outre, elles traitent avec rigueur le thème des biais algorithmiques (*bias*) et, surtout, le danger concret de discrimination pouvant résulter de l'absence de modèles de langage traduits dans toutes les langues du monde. Si l'Intelligence Artificielle est principalement entraînée en anglais et dans quelques autres langues européennes, les utilisateurs de langues moins représentées pourraient subir des performances significativement inférieures, perpétuant et amplifiant les inégalités mondiales déjà existantes.

Outre la formation et la participation internationale, le Gouvernorat continue d'investir dans un autre domaine crucial : la souveraineté et la gestion responsable des données. Ce sujet prend une importance particulière si l'on considère les données comme le « pétrole » du XXI<sup>e</sup> siècle — une ressource extraordinairement précieuse à créer, conserver, protéger et valoriser.

Pour une entité comme le Gouvernorat, le contrôle et la protection de la souveraineté des données n'est pas une question secondaire, mais une exigence incontournable. Il n'est pas acceptable que des données sensibles concernant les opérations de l'État soient gérées par des serveurs externes contrôlés par des entreprises privées étrangères. Il est nécessaire de développer des capacités internes de gestion et de protection.

La solution que le Gouvernorat entend évaluer et adopter suit cette approche : développer des solutions d'Intelligence Artificielle locales et internes. Cela implique de se doter d'un matériel informatique adéquat en termes de puissance de calcul, mais aussi respectueux des enjeux énergétiques et de la durabilité environnementale. L'objectif est de traiter les données internes de manière autonome, de les valoriser et de les rendre exploitables selon les nouveaux paradigmes de l'IA *générative*, sans dépendre d'infrastructures externes.

Cette approche présente des défis significatifs. Il s'agit de scénarios nouveaux, en grande partie inexplorés, riches en pièges techniques et organisationnels. Cependant, ils recèlent également de grandes opportunités. Ils représentent un modèle alternatif à la centralisation du pouvoir technologique entre les mains de quelques acteurs globaux. Ils démontrent qu'il est possible de développer une Intelligence Artificielle responsable, éthiquement consciente et souveraine, même pour des entités de taille plus modeste.

La technologie n'est pas neutre : les choix faits aujourd'hui détermineront le paysage de demain, et la responsabilité de gouverner l'innovation incombe à tous les acteurs de la société, des institutions aux individus.

Le Gouvernorat travaille activement pour atteindre une configuration stable, contemporaine et prête à affronter pleinement l'ère de l'Intelligence Artificielle. Ce n'est pas une tâche qui puisse être déléguée à une seule partie de l'organisation. Elle nécessite la contribution et la collaboration consciente de tous. Elle exige que chacun développe une compréhension appropriée des risques et des opportunités, et qu'il agisse avec intégrité et responsabilité éthique dans ses choix quotidiens.

L'objectif est qu'il est tout à fait possible d'adopter l'Intelligence Artificielle sans sacrifier les valeurs fondamentales de souveraineté, d'éthique, d'équité et de transparence. Une formation consciente, une participation globale et la souveraineté technologique constituent les fondations sur lesquelles bâtir l'innovation.

D. G.

# LES COMMISSIONS DU GOUVERNORAT : DES INSTRUMENTS DE GOUVERNANCE SPÉCIALISÉE

Les Commissions du Governorat ont été créées pour assister les Organes de Gouvernement dans des secteurs nécessitant une expertise particulière, tout en garantissant dans le même temps transparence, collégialité dans les décisions et conformité aux principes éthiques qui caractérisent l'action de l'État. Au fil des décennies, des commissions ont été instituées pour des matières telles que le personnel, la discipline, les questions monétaires et la sélection des collaborateurs laïcs, chacune exerçant des fonctions consultatives, délibératives ou de contrôle, selon le domaine de compétence.

C'est dans ce contexte de *gouvernance* spécialisée que s'inscrit la récente création de la Commission sur l'Intelligence Artificielle, prévue par le décret n. DCCII du 30 décembre 2024 et entrée en vigueur le 1<sup>er</sup> janvier 2025. Face à la diffusion rapide des systèmes d'Intelligence Artificielle et à leur influence croissante sur tous les aspects de la vie sociale et institutionnelle, le Governorat a jugé nécessaire de se doter d'un organisme dédié, garantissant une utilisation éthique, transparente et responsable de cette technologie.

La Commission a pour mission de préparer des lois et des règlements d'application des lignes directrices sur l'Intelligence Artificielle, de donner son avis sur les propositions d'expérimentation et d'application des systèmes d'IA, de réaliser des activités de suivi continu en signalant les

risques potentiels et de rédiger des rapports semestriels sur l'impact de l'utilisation de l'Intelligence Artificielle dans l'État de la Cité du Vatican. En substance, la Commission agit comme une boussole d'orientation pour garantir que le développement technologique reste toujours au service de la dignité humaine et du bien commun, en pleine cohérence avec les valeurs fondamentales du Saint-Siège.

La Commission est composée de cinq membres nommés par le Président du Governorat et présidée par le Secrétaire général, comme le prévoit l'article 14 du décret. Les membres proviennent de trois directions stratégiques : le Service juridique, la Direction des télécommunications et des systèmes informatiques, et la Direction des services de sécurité et de protection civile. Cette composition multidisciplinaire garantit que les décisions tiennent compte des aspects juridiques, technologiques et de sécurité. Le mandat est d'une durée de trois ans, renouvelable.

La Commission s'est déjà réunie et travaille activement à la mise en œuvre des premières étapes opérationnelles. La rapidité de l'action reflète la conscience de la délicatesse du sujet et de l'importance de disposer dès à présent d'une structure concrète, capable de répondre efficacement aux défis que l'Intelligence Artificielle pose quotidiennement à l'administration du Vatican.

D. G.



