



SALA STAMPA DELLA SANTA SEDE  
**BOLLETTINO**

HOLY SEE PRESS OFFICE BUREAU DE PRESSE DU SAINT-SIÈGE PRESSEAMT DES HEILIGEN STUHLS  
OFICINA DE PRENSA DE LA SANTA SEDE SALA DE IMPRENSA DA SANTA SÉ  
BIURO PRASOWE STOLICY APOSTOLSKIEJ دار الصحافة التابعة للكرسي الرسولي

N. 0884

Giovedì 14.12.2023

**Messaggio del Santo Padre Francesco per la 57 ma Giornata Mondiale della Pace (1° gennaio 2024)**

[Messaggio del Santo Padre](#)

[Traduzione in lingua francese](#)

[Traduzione in lingua inglese](#)

[Traduzione in lingua tedesca](#)

[Traduzione in lingua spagnola](#)

[Traduzione in lingua portoghese](#)

[Traduzione in lingua polacca](#)

[Traduzione in lingua araba](#)

Pubblichiamo di seguito il testo del Messaggio del Santo Padre Francesco per la 57ma Giornata Mondiale della Pace, che si celebra il 1° gennaio 2024 sul tema “Intelligenza artificiale e pace”:

[Messaggio del Santo Padre](#)

*Intelligenza artificiale e pace*

All’inizio del nuovo anno, tempo di grazia che il Signore dona a ciascuno di noi, vorrei rivolgermi al Popolo di Dio, alle nazioni, ai Capi di Stato e di Governo, ai Rappresentanti delle diverse religioni e della società civile, a tutti gli uomini e le donne del nostro tempo per porgere i miei auguri di pace.

## 1. Il progresso della scienza e della tecnologia come via verso la pace

La Sacra Scrittura attesta che Dio ha donato agli uomini il suo Spirito affinché abbiano «saggezza, intelligenza e scienza in ogni genere di lavoro» (Es 35,31). L'intelligenza è espressione della dignità donataci dal Creatore, che ci ha fatti a sua immagine e somiglianza (cfr Gen 1,26) e ci ha messo in grado di rispondere al suo amore attraverso la libertà e la conoscenza. La scienza e la tecnologia manifestano in modo particolare tale qualità fondamentalmente relazionale dell'intelligenza umana: sono prodotti straordinari del suo potenziale creativo.

Nella Costituzione Pastorale *Gaudium et spes*, il Concilio Vaticano II ha ribadito questa verità, dichiarando che «col suo lavoro e col suo ingegno l'uomo ha cercato sempre di sviluppare la propria vita»[1]. Quando gli esseri umani, «con l'aiuto della tecnica», si sforzano affinché la terra «diventi una dimora degna di tutta la famiglia umana»[2], agiscono secondo il disegno di Dio e cooperano con la sua volontà di portare a compimento la creazione e di diffondere la pace tra i popoli. Anche il progresso della scienza e della tecnica, nella misura in cui contribuisce a un migliore ordine della società umana, ad accrescere la libertà e la comunione fraterna, porta dunque al miglioramento dell'uomo e alla trasformazione del mondo.

Giustamente ci rallegriamo e siamo riconoscenti per le straordinarie conquiste della scienza e della tecnologia, grazie alle quali si è posto rimedio a innumerevoli mali che affliggevano la vita umana e causavano grandi sofferenze. Allo stesso tempo, i progressi tecnico-scientifici, rendendo possibile l'esercizio di un controllo finora inedito sulla realtà, stanno mettendo nelle mani dell'uomo una vasta gamma di possibilità, alcune delle quali possono rappresentare un rischio per la sopravvivenza e un pericolo per la casa comune[3].

I notevoli progressi delle nuove tecnologie dell'informazione, specialmente nella sfera digitale, presentano dunque entusiasmanti opportunità e gravi rischi, con serie implicazioni per il perseguimento della giustizia e dell'armonia tra i popoli. È pertanto necessario porsi alcune domande urgenti. Quali saranno le conseguenze, a medio e a lungo termine, delle nuove tecnologie digitali? E quale impatto avranno sulla vita degli individui e della società, sulla stabilità internazionale e sulla pace?

## 2. Il futuro dell'intelligenza artificiale tra promesse e rischi

I progressi dell'informatica e lo sviluppo delle tecnologie digitali negli ultimi decenni hanno già iniziato a produrre profonde trasformazioni nella società globale e nelle sue dinamiche. I nuovi strumenti digitali stanno cambiando il volto delle comunicazioni, della pubblica amministrazione, dell'istruzione, dei consumi, delle interazioni personali e di innumerevoli altri aspetti della vita quotidiana.

Inoltre, le tecnologie che impiegano una molteplicità di algoritmi possono estrarre, dalle tracce digitali lasciate su *internet*, dati che consentono di controllare le abitudini mentali e relazionali delle persone a fini commerciali o politici, spesso a loro insaputa, limitandone il consapevole esercizio della libertà di scelta. Infatti, in uno spazio come il *web*, caratterizzato da un sovraccarico di informazioni, possono strutturare il flusso di dati secondo criteri di selezione non sempre percepiti dall'utente.

Dobbiamo ricordare che la ricerca scientifica e le innovazioni tecnologiche non sono disincarnate dalla realtà e «neutrali»[4], ma soggette alle influenze culturali. In quanto attività pienamente umane, le direzioni che prendono riflettono scelte condizionate dai valori personali, sociali e culturali di ogni epoca. Dicasi lo stesso per i risultati che conseguono: essi, proprio in quanto frutto di approcci specificamente umani al mondo circostante, hanno sempre una dimensione etica, strettamente legata alle decisioni di chi progetta la sperimentazione e indirizza la produzione verso particolari obiettivi.

Questo vale anche per le forme di intelligenza artificiale. Di essa, ad oggi, non esiste una definizione univoca nel mondo della scienza e della tecnologia. Il termine stesso, ormai entrato nel linguaggio comune, abbraccia una varietà di scienze, teorie e tecniche volte a far sì che le macchine riproducano o imitino, nel loro funzionamento, le capacità cognitive degli esseri umani. Parlare al plurale di "forme di intelligenza" può aiutare a sottolineare soprattutto il divario incolmabile che esiste tra questi sistemi, per quanto sorprendenti e potenti, e la persona umana: essi sono, in ultima analisi, "frammentari", nel senso che possono solo imitare o riprodurre alcune

funzioni dell'intelligenza umana. L'uso del plurale evidenzia inoltre che questi dispositivi, molto diversi tra loro, vanno sempre considerati come "sistemi socio-tecnici". Infatti il loro impatto, al di là della tecnologia di base, dipende non solo dalla progettazione, ma anche dagli obiettivi e dagli interessi di chi li possiede e di chi li sviluppa, nonché dalle situazioni in cui vengono impiegati.

L'intelligenza artificiale, quindi, deve essere intesa come una galassia di realtà diverse e non possiamo presumere a priori che il suo sviluppo apporti un contributo benefico al futuro dell'umanità e alla pace tra i popoli. Tale risultato positivo sarà possibile solo se ci dimostreremo capaci di agire in modo responsabile e di rispettare valori umani fondamentali come «l'inclusione, la trasparenza, la sicurezza, l'equità, la riservatezza e l'affidabilità»[5].

Non è sufficiente nemmeno presumere, da parte di chi progetta algoritmi e tecnologie digitali, un impegno ad agire in modo etico e responsabile. Occorre rafforzare o, se necessario, istituire organismi incaricati di esaminare le questioni etiche emergenti e di tutelare i diritti di quanti utilizzano forme di intelligenza artificiale o ne sono influenzati[6].

L'immensa espansione della tecnologia deve quindi essere accompagnata da un'adeguata formazione alla responsabilità per il suo sviluppo. La libertà e la convivenza pacifica sono minacciate quando gli esseri umani cedono alla tentazione dell'egoismo, dell'interesse personale, della brama di profitto e della sete di potere. Abbiamo perciò il dovere di allargare lo sguardo e di orientare la ricerca tecnico-scientifica al perseguimento della pace e del bene comune, al servizio dello sviluppo integrale dell'uomo e della comunità[7].

La dignità intrinseca di ogni persona e la fraternità che ci lega come membri dell'unica famiglia umana devono stare alla base dello sviluppo di nuove tecnologie e servire come criteri indiscutibili per valutarle prima del loro impiego, in modo che il progresso digitale possa avvenire nel rispetto della giustizia e contribuire alla causa della pace. Gli sviluppi tecnologici che non portano a un miglioramento della qualità di vita di tutta l'umanità, ma al contrario aggravano le disuguaglianze e i conflitti, non potranno mai essere considerati vero progresso[8].

L'intelligenza artificiale diventerà sempre più importante. Le sfide che pone sono tecniche, ma anche antropologiche, educative, sociali e politiche. Promette, ad esempio, un risparmio di fatiche, una produzione più efficiente, trasporti più agevoli e mercati più dinamici, oltre a una rivoluzione nei processi di raccolta, organizzazione e verifica dei dati. Occorre essere consapevoli delle rapide trasformazioni in atto e gestirle in modo da salvaguardare i diritti umani fondamentali, rispettando le istituzioni e le leggi che promuovono lo sviluppo umano integrale. L'intelligenza artificiale dovrebbe essere al servizio del migliore potenziale umano e delle nostre più alte aspirazioni, non in competizione con essi.

### 3. La tecnologia del futuro: macchine che imparano da sole

Nelle sue molteplici forme l'intelligenza artificiale, basata su tecniche di apprendimento automatico (*machine learning*), pur essendo ancora in fase pionieristica, sta già introducendo notevoli cambiamenti nel tessuto delle società, esercitando una profonda influenza sulle culture, sui comportamenti sociali e sulla costruzione della pace.

Sviluppi come il *machine learning* o come l'apprendimento profondo (*deep learning*) sollevano questioni che trascendono gli ambiti della tecnologia e dell'ingegneria e hanno a che fare con una comprensione strettamente connessa al significato della vita umana, ai processi basilari della conoscenza e alla capacità della mente di raggiungere la verità.

L'abilità di alcuni dispositivi nel produrre testi sintatticamente e semanticamente coerenti, ad esempio, non è garanzia di affidabilità. Si dice che possano "allucinare", cioè generare affermazioni che a prima vista sembrano plausibili, ma che in realtà sono infondate o tradiscono pregiudizi. Questo pone un serio problema quando l'intelligenza artificiale viene impiegata in campagne di disinformazione che diffondono notizie false e portano a una crescente sfiducia nei confronti dei mezzi di comunicazione. La riservatezza, il possesso dei dati e la proprietà intellettuale sono altri ambiti in cui le tecnologie in questione comportano gravi rischi, a cui si

aggiungono ulteriori conseguenze negative legate a un loro uso improprio, come la discriminazione, l'interferenza nei processi elettorali, il prendere piede di una società che sorveglia e controlla le persone, l'esclusione digitale e l'inasprimento di un individualismo sempre più scollegato dalla collettività. Tutti questi fattori rischiano di alimentare i conflitti e di ostacolare la pace.

#### 4. *Il senso del limite nel paradigma tecnocratico*

Il nostro mondo è troppo vasto, vario e complesso per essere completamente conosciuto e classificato. La mente umana non potrà mai esaurirne la ricchezza, nemmeno con l'aiuto degli algoritmi più avanzati. Questi, infatti, non offrono previsioni garantite del futuro, ma solo approssimazioni statistiche. Non tutto può essere pronosticato, non tutto può essere calcolato; alla fine «la realtà è superiore all'idea»[9] e, per quanto prodigiosa possa essere la nostra capacità di calcolo, ci sarà sempre un residuo inaccessibile che sfugge a qualsiasi tentativo di misurazione.

Inoltre, la grande quantità di dati analizzati dalle intelligenze artificiali non è di per sé garanzia di imparzialità. Quando gli algoritmi estrapolano informazioni, corrono sempre il rischio di distorcerle, replicando le ingiustizie e i pregiudizi degli ambienti in cui esse hanno origine. Più diventano veloci e complessi, più è difficile comprendere perché abbiano prodotto un determinato risultato.

Le macchine "intelligenti" possono svolgere i compiti loro assegnati con sempre maggiore efficienza, ma lo scopo e il significato delle loro operazioni continueranno a essere determinati o abilitati da esseri umani in possesso di un proprio universo di valori. Il rischio è che i criteri alla base di certe scelte diventino meno chiari, che la responsabilità decisionale venga nascosta e che i produttori possano sottrarsi all'obbligo di agire per il bene della comunità. In un certo senso, ciò è favorito dal sistema tecnocratico, che allea l'economia con la tecnologia e privilegia il criterio dell'efficienza, tendendo a ignorare tutto ciò che non è legato ai suoi interessi immediati[10].

Questo deve farci riflettere su un aspetto tanto spesso trascurato nella mentalità attuale, tecnocratica ed efficientista, quanto decisivo per lo sviluppo personale e sociale: il "senso del limite". L'essere umano, infatti, mortale per definizione, pensando di travalicare ogni limite in virtù della tecnica, rischia, nell'ossessione di voler controllare tutto, di perdere il controllo su sé stesso; nella ricerca di una libertà assoluta, di cadere nella spirale di una dittatura tecnologica. Riconoscere e accettare il proprio limite di creatura è per l'uomo condizione indispensabile per conseguire, o meglio, accogliere in dono la pienezza. Invece, nel contesto ideologico di un paradigma tecnocratico, animato da una prometeica presunzione di autosufficienza, le disuguaglianze potrebbero crescere a dismisura, e la conoscenza e la ricchezza accumularsi nelle mani di pochi, con gravi rischi per le società democratiche e la coesistenza pacifica[11].

#### 5. *Temi scottanti per l'etica*

In futuro, l'affidabilità di chi richiede un mutuo, l'idoneità di un individuo ad un lavoro, la possibilità di recidiva di un condannato o il diritto a ricevere asilo politico o assistenza sociale potrebbero essere determinati da sistemi di intelligenza artificiale. La mancanza di diversificati livelli di mediazione che questi sistemi introducono è particolarmente esposta a forme di pregiudizio e discriminazione: gli errori sistemici possono facilmente moltiplicarsi, producendo non solo ingiustizie in singoli casi ma anche, per effetto domino, vere e proprie forme di disuguaglianza sociale.

Talvolta, inoltre, le forme di intelligenza artificiale sembrano in grado di influenzare le decisioni degli individui attraverso opzioni predeterminate associate a stimoli e dissuasioni, oppure mediante sistemi di regolazione delle scelte personali basati sull'organizzazione delle informazioni. Queste forme di manipolazione o di controllo sociale richiedono un'attenzione e una supervisione accurate, e implicano una chiara responsabilità legale da parte dei produttori, di chi le impiega e delle autorità governative.

L'affidamento a processi automatici che categorizzano gli individui, ad esempio attraverso l'uso pervasivo della vigilanza o l'adozione di sistemi di credito sociale, potrebbe avere ripercussioni profonde anche sul tessuto

civile, stabilendo improprie graduatorie tra i cittadini. E questi processi artificiali di classificazione potrebbero portare anche a conflitti di potere, non riguardando solo destinatari virtuali, ma persone in carne ed ossa. Il rispetto fondamentale per la dignità umana postula di rifiutare che l'unicità della persona venga identificata con un insieme di dati. Non si deve permettere agli algoritmi di determinare il modo in cui intendiamo i diritti umani, di mettere da parte i valori essenziali della compassione, della misericordia e del perdono o di eliminare la possibilità che un individuo cambi e si lasci alle spalle il passato.

In questo contesto non possiamo fare a meno di considerare l'impatto delle nuove tecnologie in ambito lavorativo: mansioni che un tempo erano appannaggio esclusivo della manodopera umana vengono rapidamente assorbite dalle applicazioni industriali dell'intelligenza artificiale. Anche in questo caso, c'è il rischio sostanziale di un vantaggio sproporzionato per pochi a scapito dell'impoverimento di molti. Il rispetto della dignità dei lavoratori e l'importanza dell'occupazione per il benessere economico delle persone, delle famiglie e delle società, la sicurezza degli impieghi e l'equità dei salari dovrebbero costituire un'alta priorità per la Comunità internazionale, mentre queste forme di tecnologia penetrano sempre più profondamente nei luoghi di lavoro.

## 6. *Trasformeremo le spade in vomeri?*

In questi giorni, guardando il mondo che ci circonda, non si può sfuggire alle gravi questioni etiche legate al settore degli armamenti. La possibilità di condurre operazioni militari attraverso sistemi di controllo remoto ha portato a una minore percezione della devastazione da essi causata e della responsabilità del loro utilizzo, contribuendo a un approccio ancora più freddo e distaccato all'immensa tragedia della guerra. La ricerca sulle tecnologie emergenti nel settore dei cosiddetti "sistemi d'arma autonomi letali", incluso l'utilizzo bellico dell'intelligenza artificiale, è un grave motivo di preoccupazione etica. I sistemi d'arma autonomi non potranno mai essere soggetti moralmente responsabili: l'esclusiva capacità umana di giudizio morale e di decisione etica è più di un complesso insieme di algoritmi, e tale capacità non può essere ridotta alla programmazione di una macchina che, per quanto "intelligente", rimane pur sempre una macchina. Per questo motivo, è imperativo garantire una supervisione umana adeguata, significativa e coerente dei sistemi d'arma.

Non possiamo nemmeno ignorare la possibilità che armi sofisticate finiscano nelle mani sbagliate, facilitando, ad esempio, attacchi terroristici o interventi volti a destabilizzare istituzioni di governo legittime. Il mondo, insomma, non ha proprio bisogno che le nuove tecnologie contribuiscano all'iniquo sviluppo del mercato e del commercio delle armi, promuovendo la follia della guerra. Così facendo, non solo l'intelligenza, ma il cuore stesso dell'uomo, correrà il rischio di diventare sempre più "artificiale". Le più avanzate applicazioni tecniche non vanno impiegate per agevolare la risoluzione violenta dei conflitti, ma per pavimentare le vie della pace.

In un'ottica più positiva, se l'intelligenza artificiale fosse utilizzata per promuovere lo sviluppo umano integrale, potrebbe introdurre importanti innovazioni nell'agricoltura, nell'istruzione e nella cultura, un miglioramento del livello di vita di intere nazioni e popoli, la crescita della fraternità umana e dell'amicizia sociale. In definitiva, il modo in cui la utilizziamo per includere gli ultimi, cioè i fratelli e le sorelle più deboli e bisognosi, è la misura rivelatrice della nostra umanità.

Uno sguardo umano e il desiderio di un futuro migliore per il nostro mondo portano alla necessità di un dialogo interdisciplinare finalizzato a uno sviluppo etico degli algoritmi – *l'algor-etica* –, in cui siano i valori a orientare i percorsi delle nuove tecnologie[12]. Le questioni etiche dovrebbero essere tenute in considerazione fin dall'inizio della ricerca, così come nelle fasi di sperimentazione, progettazione, produzione, distribuzione e commercializzazione. Questo è l'approccio dell'etica della progettazione, in cui le istituzioni educative e i responsabili del processo decisionale hanno un ruolo essenziale da svolgere.

## 7. *Sfide per l'educazione*

Lo sviluppo di una tecnologia che rispetti e serva la dignità umana ha chiare implicazioni per le istituzioni educative e per il mondo della cultura. Moltiplicando le possibilità di comunicazione, le tecnologie digitali hanno permesso di incontrarsi in modi nuovi. Tuttavia, rimane la necessità di una riflessione continua sul tipo di

relazioni a cui ci stanno indirizzando. I giovani stanno crescendo in ambienti culturali pervasi dalla tecnologia e questo non può non mettere in discussione i metodi di insegnamento e formazione.

L'educazione all'uso di forme di intelligenza artificiale dovrebbe mirare soprattutto a promuovere il pensiero critico. È necessario che gli utenti di ogni età, ma soprattutto i giovani, sviluppino una capacità di discernimento nell'uso di dati e contenuti raccolti sul *web* o prodotti da sistemi di intelligenza artificiale. Le scuole, le università e le società scientifiche sono chiamate ad aiutare gli studenti e i professionisti a fare propri gli aspetti sociali ed etici dello sviluppo e dell'utilizzo della tecnologia.

La formazione all'uso dei nuovi strumenti di comunicazione dovrebbe tenere conto non solo della disinformazione, delle *fake news*, ma anche dell'inquietante recrudescenza di «paure ancestrali [...] che hanno saputo nascondersi e potenziarsi dietro nuove tecnologie»[13]. Purtroppo, ancora una volta ci troviamo a dover combattere «la tentazione di fare una cultura dei muri, di alzare muri per impedire l'incontro con altre culture, con altra gente»[14] e lo sviluppo di una coesistenza pacifica e fraterna.

#### 8. *Sfide per lo sviluppo del diritto internazionale*

La portata globale dell'intelligenza artificiale rende evidente che, accanto alla responsabilità degli Stati sovrani di disciplinarne l'uso al proprio interno, le Organizzazioni internazionali possono svolgere un ruolo decisivo nel raggiungere accordi multilaterali e nel coordinarne l'applicazione e l'attuazione[15]. A tale proposito, esorto la Comunità delle nazioni a lavorare unita al fine di adottare un trattato internazionale vincolante, che regoli lo sviluppo e l'uso dell'intelligenza artificiale nelle sue molteplici forme. L'obiettivo della regolamentazione, naturalmente, non dovrebbe essere solo la prevenzione delle cattive pratiche, ma anche l'incoraggiamento delle buone pratiche, stimolando approcci nuovi e creativi e facilitando iniziative personali e collettive[16].

In definitiva, nella ricerca di modelli normativi che possano fornire una guida etica agli sviluppatori di tecnologie digitali, è indispensabile identificare i valori umani che dovrebbero essere alla base dell'impegno delle società per formulare, adottare e applicare necessari quadri legislativi. Il lavoro di redazione di linee guida etiche per la produzione di forme di intelligenza artificiale non può prescindere dalla considerazione di questioni più profonde riguardanti il significato dell'esistenza umana, la tutela dei diritti umani fondamentali, il perseguimento della giustizia e della pace. Questo processo di discernimento etico e giuridico può rivelarsi un'occasione preziosa per una riflessione condivisa sul ruolo che la tecnologia dovrebbe avere nella nostra vita individuale e comunitaria e su come il suo utilizzo possa contribuire alla creazione di un mondo più equo e umano. Per questo motivo, nei dibattiti sulla regolamentazione dell'intelligenza artificiale, si dovrebbe tenere conto della voce di tutte le parti interessate, compresi i poveri, gli emarginati e altri che spesso rimangono inascoltati nei processi decisionali globali.

\* \* \* \* \*

Spero che questa riflessione incoraggi a far sì che i progressi nello sviluppo di forme di intelligenza artificiale servano, in ultima analisi, la causa della fraternità umana e della pace. Non è responsabilità di pochi, ma dell'intera famiglia umana. La pace, infatti, è il frutto di relazioni che riconoscono e accolgono l'altro nella sua inalienabile dignità, e di cooperazione e impegno nella ricerca dello sviluppo integrale di tutte le persone e di tutti i popoli.

La mia preghiera all'inizio del nuovo anno è che il rapido sviluppo di forme di intelligenza artificiale non accresca le troppe disuguaglianze e ingiustizie già presenti nel mondo, ma contribuisca a porre fine a guerre e conflitti, e ad alleviare molte forme di sofferenza che affliggono la famiglia umana. Possano i fedeli cristiani, i credenti di varie religioni e gli uomini e le donne di buona volontà collaborare in armonia per cogliere le opportunità e affrontare le sfide poste dalla rivoluzione digitale, e consegnare alle generazioni future un mondo più solidale, giusto e pacifico.

Dal Vaticano, 8 dicembre 2023

---

[1] N. 33.

[2] *Ibid.*, 57.

[3] Cfr Lett. enc. *Laudato si'* (24 maggio 2015), 104.

[4] Cfr *ibid.*, 114.

[5] *Udienza ai partecipanti all'Incontro "Minerva Dialogues"* (27 marzo 2023).

[6] Cfr *ibid.*

[7] Cfr *Messaggio al Presidente Esecutivo del "World Economic Forum" a Davos-Klosters* (12 gennaio 2018).

[8] Cfr Lett. enc. *Laudato si'*, 194; *Discorso ai partecipanti al Seminario "Il bene comune nell'era digitale"* (27 settembre 2019).

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

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

[11] Cfr *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita* (28 febbraio 2020).

[12] Cfr *ibid.*

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

[14] Cfr *ibid.*

[15] Cfr *ibid.*, 170-175.

[16] Cfr Lett. enc. *Laudato si'*, 177.

[01927-IT.01] [Testo originale: Italiano]

### **Traduzione in lingua francese**

#### ***Intelligence artificielle et paix***

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 étonnants et puissants 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 lie 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 courent 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 privilégie 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 paver 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 régleme 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 affligent 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.

[01927-FR.01] [Texte original: Italien]

## Traduzione in lingua inglese

### *Artificial Intelligence and Peace*

At the beginning of the New Year, a time of grace which the Lord gives to each one of us, I would like to address God's People, the various nations, heads of state and government, the leaders of the different religions and civil society, and all the men and women of our time, in order to offer my fervent good wishes for peace.

#### *1. The progress of science and technology as a path to peace*

Sacred Scripture attests that God bestowed his Spirit upon human beings so that they might have "skill and understanding and knowledge in every craft" (*Ex 35:31*). Human intelligence is an expression of the dignity with which we have been endowed by the Creator, who made us in his own image and likeness (cf. *Gen 1:26*), and enabled us to respond consciously and freely to his love. In a particular way, science and technology manifest this fundamentally relational quality of human intelligence; they are brilliant products of its creative potential.

In its Pastoral Constitution *Gaudium et Spes*, the Second Vatican Council restated this truth, declaring that "through its labours and its native endowments, humanity has ceaselessly sought to better its life".[1] When human beings, "with the aid of technology", endeavour to make "the earth a dwelling worthy of the whole human family",[2] they carry out God's plan and cooperate with his will to perfect creation and bring about peace among peoples. Progress in science and technology, insofar as it contributes to greater order in human society and greater fraternal communion and freedom, thus leads to the betterment of humanity and the transformation of the world.

We rightly rejoice and give thanks for the impressive achievements of science and technology, as a result of which countless ills that formerly plagued human life and caused great suffering have been remedied. At the same time, techno-scientific advances, by making it possible to exercise hitherto unprecedented control over reality, are placing in human hands a vast array of options, including some that may pose a risk to our survival and endanger our common home.[3]

The remarkable advances in new information technologies, particularly in the digital sphere, thus offer exciting opportunities and grave risks, with serious implications for the pursuit of justice and harmony among peoples. Any number of urgent questions need to be asked. What will be the consequences, in the medium and long term, of these new digital technologies? And what impact will they have on individual lives and on societies, on international stability and peace?

#### *2. The future of artificial intelligence: between promise and risk*

Progress in information technology and the development of digital technologies in recent decades have already begun to effect profound transformations in global society and its various dynamics. New digital tools are even now changing the face of communications, public administration, education, consumption, personal interactions and countless other aspects of our daily lives.

Moreover, from the digital footprints spread throughout the Internet, technologies employing a variety of algorithms can extract data that enable them to control mental and relational habits for commercial or political purposes, often without our knowledge, thus limiting our conscious exercise of freedom of choice. In a space like the Web, marked by information overload, they can structure the flow of data according to criteria of selection that are not always perceived by the user.

We need to remember that scientific research and technological innovations are not disembodied and “neutral”, [4] but subject to cultural influences. As fully human activities, the directions they take reflect choices conditioned by personal, social and cultural values in any given age. The same must be said of the results they produce: precisely as the fruit of specifically human ways of approaching the world around us, the latter always have an ethical dimension, closely linked to decisions made by those who design their experimentation and direct their production towards particular objectives.

This is also the case with forms of artificial intelligence. To date, there is no single definition of artificial intelligence in the world of science and technology. The term itself, which by now has entered into everyday parlance, embraces a variety of sciences, theories and techniques aimed at making machines reproduce or imitate in their functioning the cognitive abilities of human beings. To speak in the plural of “forms of intelligence” can help to emphasize above all the unbridgeable gap between such systems, however amazing and powerful, and the human person: in the end, they are merely “fragmentary”, in the sense that they can only imitate or reproduce certain functions of human intelligence. The use of the plural likewise brings out the fact that these devices greatly differ among themselves and that they should always be regarded as “socio-technical systems”. For the impact of any artificial intelligence device – regardless of its underlying technology – depends not only on its technical design, but also on the aims and interests of its owners and developers, and on the situations in which it will be employed.

Artificial intelligence, then, ought to be understood as a galaxy of different realities. We cannot presume a priori that its development will make a beneficial contribution to the future of humanity and to peace among peoples. That positive outcome will only be achieved if we show ourselves capable of acting responsibly and respect such fundamental human values as “inclusion, transparency, security, equity, privacy and reliability”. [5]

Nor is it sufficient simply to presume a commitment on the part of those who design algorithms and digital technologies to act ethically and responsibly. There is a need to strengthen or, if necessary, to establish bodies charged with examining the ethical issues arising in this field and protecting the rights of those who employ forms of artificial intelligence or are affected by them. [6]

The immense expansion of technology thus needs to be accompanied by an appropriate formation in responsibility for its future development. Freedom and peaceful coexistence are threatened whenever human beings yield to the temptation to selfishness, self-interest, the desire for profit and the thirst for power. We thus have a duty to broaden our gaze and to direct techno-scientific research towards the pursuit of peace and the common good, in the service of the integral development of individuals and communities. [7]

The inherent dignity of each human being and the fraternity that binds us together as members of the one human family must undergird the development of new technologies and serve as indisputable criteria for evaluating them before they are employed, so that digital progress can occur with due respect for justice and contribute to the cause of peace. Technological developments that do not lead to an improvement in the quality of life of all humanity, but on the contrary aggravate inequalities and conflicts, can never count as true progress. [8]

Artificial intelligence will become increasingly important. The challenges it poses are technical, but also anthropological, educational, social and political. It promises, for instance, liberation from drudgery, more efficient manufacturing, easier transport and more ready markets, as well as a revolution in processes of accumulating, organizing and confirming data. We need to be aware of the rapid transformations now taking place and to manage them in ways that safeguard fundamental human rights and respect the institutions and laws that promote integral human development. Artificial intelligence ought to serve our best human potential and our highest aspirations, not compete with them.

### *3. The technology of the future: machines that “learn” by themselves*

In its multiple forms, artificial intelligence based on machine learning techniques, while still in its pioneering phases, is already introducing considerable changes to the fabric of societies and exerting a profound influence

on cultures, societal behaviours and peacebuilding.

Developments such as machine learning or deep learning, raise questions that transcend the realms of technology and engineering, and have to do with the deeper understanding of the meaning of human life, the construction of knowledge, and the capacity of the mind to attain truth.

The ability of certain devices to produce syntactically and semantically coherent texts, for example, is no guarantee of their reliability. They are said to “hallucinate”, that is, to create statements that at first glance appear plausible but are unfounded or betray biases. This poses a serious problem when artificial intelligence is deployed in campaigns of disinformation that spread false news and lead to a growing distrust of the communications media. Privacy, data ownership and intellectual property are other areas where these technologies engender grave risks. To which we can add other negative consequences of the misuse of these technologies, such as discrimination, interference in elections, the rise of a surveillance society, digital exclusion and the exacerbation of an individualism increasingly disconnected from society. All these factors risk fueling conflicts and hindering peace.

#### 4. *The sense of limit in the technocratic paradigm*

Our world is too vast, varied and complex ever to be fully known and categorized. The human mind can never exhaust its richness, even with the aid of the most advanced algorithms. Such algorithms do not offer guaranteed predictions of the future, but only statistical approximations. Not everything can be predicted, not everything can be calculated; in the end, “realities are greater than ideas”.<sup>[9]</sup> No matter how prodigious our calculating power may be, there will always be an inaccessible residue that evades any attempt at quantification.

In addition, the vast amount of data analyzed by artificial intelligences is in itself no guarantee of impartiality. When algorithms extrapolate information, they always run the risk of distortion, replicating the injustices and prejudices of the environments where they originate. The faster and more complex they become, the more difficult it proves to understand why they produced a particular result.

“Intelligent” machines may perform the tasks assigned to them with ever greater efficiency, but the purpose and the meaning of their operations will continue to be determined or enabled by human beings possessed of their own universe of values. There is a risk that the criteria behind certain decisions will become less clear, responsibility for those decisions concealed, and producers enabled to evade their obligation to act for the benefit of the community. In some sense, this is favoured by the technocratic system, which allies the economy with technology and privileges the criterion of efficiency, tending to ignore anything unrelated to its immediate interests.<sup>[10]</sup>

This should lead us to reflect on something frequently overlooked in our current technocratic and efficiency-oriented mentality, as it is decisive for personal and social development: the “sense of limit”. Human beings are, by definition, mortal; by proposing to overcome every limit through technology, in an obsessive desire to control everything, we risk losing control over ourselves; in the quest for an absolute freedom, we risk falling into the spiral of a “technological dictatorship”. Recognizing and accepting our limits as creatures is an indispensable condition for reaching, or better, welcoming fulfilment as a gift. In the ideological context of a technocratic paradigm inspired by a Promethean presumption of self-sufficiency, inequalities could grow out of proportion, knowledge and wealth accumulate in the hands of a few, and grave risks ensue for democratic societies and peaceful coexistence.<sup>[11]</sup>

#### 5. *Burning issues for ethics*

In the future, the reliability of an applicant for a mortgage, the suitability of an individual for a job, the possibility of recidivism on the part of a convicted person, or the right to receive political asylum or social assistance could be determined by artificial intelligence systems. The lack of different levels of mediation that these systems introduce is particularly exposed to forms of bias and discrimination: systemic errors can easily multiply, producing not only injustices in individual cases but also, due to the domino effect, real forms of social inequality.

At times too, forms of artificial intelligence seem capable of influencing individuals' decisions by operating through pre-determined options associated with stimuli and dissuasions, or by operating through a system of regulating people's choices based on information design. These forms of manipulation or social control require careful attention and oversight, and imply a clear legal responsibility on the part of their producers, their deployers, and government authorities.

Reliance on automatic processes that categorize individuals, for instance, by the pervasive use of surveillance or the adoption of social credit systems, could likewise have profound repercussions on the social fabric by establishing a ranking among citizens. These artificial processes of categorization could lead also to power conflicts, since they concern not only virtual users but real people. Fundamental respect for human dignity demands that we refuse to allow the uniqueness of the person to be identified with a set of data. Algorithms must not be allowed to determine how we understand human rights, to set aside the essential human values of compassion, mercy and forgiveness, or to eliminate the possibility of an individual changing and leaving his or her past behind.

Nor can we fail to consider, in this context, the impact of new technologies on the workplace. Jobs that were once the sole domain of human labour are rapidly being taken over by industrial applications of artificial intelligence. Here too, there is the substantial risk of disproportionate benefit for the few at the price of the impoverishment of many. Respect for the dignity of labourers and the importance of employment for the economic well-being of individuals, families, and societies, for job security and just wages, ought to be a high priority for the international community as these forms of technology penetrate more deeply into our workplaces.

#### 6. *Shall we turn swords into ploughshares?*

In these days, as we look at the world around us, there can be no escaping serious ethical questions related to the armaments sector. The ability to conduct military operations through remote control systems has led to a lessened perception of the devastation caused by those weapon systems and the burden of responsibility for their use, resulting in an even more cold and detached approach to the immense tragedy of war. Research on emerging technologies in the area of so-called Lethal Autonomous Weapon Systems, including the weaponization of artificial intelligence, is a cause for grave ethical concern. Autonomous weapon systems can never be morally responsible subjects. The unique human capacity for moral judgment and ethical decision-making is more than a complex collection of algorithms, and that capacity cannot be reduced to programming a machine, which as "intelligent" as it may be, remains a machine. For this reason, it is imperative to ensure adequate, meaningful and consistent human oversight of weapon systems.

Nor can we ignore the possibility of sophisticated weapons ending up in the wrong hands, facilitating, for instance, terrorist attacks or interventions aimed at destabilizing the institutions of legitimate systems of government. In a word, the world has no need of new technologies that contribute to the unjust development of commerce and the weapons trade and consequently end up promoting the folly of war. By so doing, not only intelligence but the human heart itself would risk becoming ever more "artificial". The most advanced technological applications should not be employed to facilitate the violent resolution of conflicts, but rather to pave the way for peace.

On a more positive note, if artificial intelligence were used to promote integral human development, it could introduce important innovations in agriculture, education and culture, an improved level of life for entire nations and peoples, and the growth of human fraternity and social friendship. In the end, the way we use it to include the least of our brothers and sisters, the vulnerable and those most in need, will be the true measure of our humanity.

An authentically humane outlook and the desire for a better future for our world surely indicates the need for a cross-disciplinary dialogue aimed at an ethical development of algorithms – an algor-ethics – in which values will shape the directions taken by new technologies.[12] Ethical considerations should also be taken into account from the very beginning of research, and continue through the phases of experimentation, design, production, distribution and marketing. This is the approach of ethics by design, and it is one in which educational institutions

and decision-makers have an essential role to play.

### *7. Challenges for education*

The development of a technology that respects and serves human dignity has clear ramifications for our educational institutions and the world of culture. By multiplying the possibilities of communication, digital technologies have allowed us to encounter one another in new ways. Yet there remains a need for sustained reflection on the kinds of relationships to which they are steering us. Our young people are growing up in cultural environments pervaded by technology, and this cannot but challenge our methods of teaching, education and training.

Education in the use of forms of artificial intelligence should aim above all at promoting critical thinking. Users of all ages, but especially the young, need to develop a discerning approach to the use of data and content collected on the web or produced by artificial intelligence systems. Schools, universities and scientific societies are challenged to help students and professionals to grasp the social and ethical aspects of the development and uses of technology.

Training in the use of new means of communication should also take account not only of disinformation, “fake news”, but also the disturbing recrudescence of “certain ancestral fears... that have been able to hide and spread behind new technologies”.<sup>[13]</sup> Sadly, we once more find ourselves having to combat “the temptation to build a culture of walls, to raise walls... in order to prevent an encounter with other cultures and other peoples”,<sup>[14]</sup> and the development of a peaceful and fraternal coexistence.

### *8. Challenges for the development of international law*

The global scale of artificial intelligence makes it clear that, alongside the responsibility of sovereign states to regulate its use internally, international organizations can play a decisive role in reaching multilateral agreements and coordinating their application and enforcement.<sup>[15]</sup> In this regard, I urge the global community of nations to work together in order to adopt a binding international treaty that regulates the development and use of artificial intelligence in its many forms. The goal of regulation, naturally, should not only be the prevention of harmful practices but also the encouragement of best practices, by stimulating new and creative approaches and encouraging individual or group initiatives.<sup>[16]</sup>

In the quest for normative models that can provide ethical guidance to developers of digital technologies, it is indispensable to identify the human values that should undergird the efforts of societies to formulate, adopt and enforce much-needed regulatory frameworks. The work of drafting ethical guidelines for producing forms of artificial intelligence can hardly prescind from the consideration of deeper issues regarding the meaning of human existence, the protection of fundamental human rights and the pursuit of justice and peace. This process of ethical and juridical discernment can prove a precious opportunity for shared reflection on the role that technology should play in our individual and communal lives, and how its use can contribute to the creation of a more equitable and humane world. For this reason, in debates about the regulation of artificial intelligence, the voices of all stakeholders should be taken into account, including the poor, the powerless and others who often go unheard in global decision-making processes.

\* \* \* \* \*

I hope that the foregoing reflection will encourage efforts to ensure that progress in developing forms of artificial intelligence will ultimately serve the cause of human fraternity and peace. It is not the responsibility of a few but of the entire human family. For peace is the fruit of relationships that recognize and welcome others in their inalienable dignity, and of cooperation and commitment in seeking the integral development of all individuals and peoples.

It is my prayer at the start of the New Year that the rapid development of forms of artificial intelligence will not

increase cases of inequality and injustice all too present in today's world, but will help put an end to wars and conflicts, and alleviate many forms of suffering that afflict our human family. May Christian believers, followers of various religions and men and women of good will work together in harmony to embrace the opportunities and confront the challenges posed by the digital revolution and thus hand on to future generations a world of greater solidarity, justice and peace.

From the Vatican, 8 December 2023

FRANCIS

---

[1] No. 33.

[2] Ibid., 57.

[3] Cf. Encyclical Letter *Laudato si'* (24 May 2015), 104.

[4] Cf. *ibid.*, 114.

[5] *Address to Participants in the "Minerva Dialogues"* (27 March 2023).

[6] Cf. *ibid.*

[7] Cf. *Message to the Executive Chairman of the "World Economic Forum" meeting in Davos* (12 January 2018).

[8] Cf. Encyclical Letter *Laudato si'* (24 May 2015), 194; *Address to Participants in the Seminar "The Common Good in the Digital Age"* (27 September 2019).

[9] Apostolic Exhortation *Evangelii Gaudium* (24 November 2013), 233.

[10] Cf. Encyclical Letter *Laudato si'* (24 May 2015), 54.

[11] Cf. *Meeting with Participants in the Plenary Assembly of the Pontifical Academy for Life* (28 February 2020).

[12] Cf. *ibid.*

[13] Encyclical Letter *Fratelli Tutti* (3 October 2020), 27.

[14] *Ibid.*

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

[16] Cf. Encyclical Letter *Laudato si'* (24 May 2015), 177.

[01927-EN.01] [Original text: Italian]

**Traduzione in lingua tedesca**

## *Künstliche Intelligenz und Frieden*

Zu Beginn des neuen Jahres, einer Zeit der Gnade, die der Herr jedem von uns gewährt, möchte ich mich an das Volk Gottes, an die Nationen, an die Staats- und Regierungschefs, an die Vertreter der verschiedenen Religionen und der Zivilgesellschaft sowie an alle Männer und Frauen unserer Zeit wenden, um ihnen meine besten Wünsche für den Frieden zu übermitteln.

### *1. Der Fortschritt von Wissenschaft und Technik als Weg zum Frieden*

Die Heilige Schrift bezeugt, dass Gott den Menschen seinen Geist gegeben hat, damit sie »mit Weisheit, Klugheit und Kenntnis für jegliche Arbeit« ausgestattet seien (*Ex 35,31*). Die Intelligenz ist Ausdruck der Würde, die uns der Schöpfer verliehen hat, der uns nach seinem Bild und Gleichnis geschaffen hat (vgl. *Gen 1,26*) und uns befähigt hat, auf seine Liebe frei und bewusst zu antworten. Wissenschaft und Technik verdeutlichen in besonderer Weise eine solche grundlegend relationale Beschaffenheit der menschlichen Intelligenz: Sie sind außergewöhnliche Ergebnisse ihres schöpferischen Potentials.

In der Pastoralconstitution *Gaudium et Spes* hat das Zweite Vatikanische Konzil diese Wahrheit bekräftigt, indem es erklärte: »Durch Arbeit und Geisteskraft hat der Mensch immer versucht, sein Leben reicher zu entfalten«[1]. Wenn die Menschen sich »mit Hilfe der Technik« darum bemühen, dass die Erde »eine würdige Wohnstätte für die gesamte menschliche Familie werde«[2], dann handeln sie nach dem Plan Gottes und arbeiten mit seinem Willen zusammen, um die Schöpfung zu vollenden und den Frieden unter den Völkern zu verbreiten. Auch der Fortschritt von Wissenschaft und Technik, soweit er zu einer besseren Ordnung der menschlichen Gesellschaft, zu wachsender Freiheit und geschwisterlicher Gemeinschaft beiträgt, führt also zur Besserung des Menschen und zur Umgestaltung der Welt.

Wir freuen uns zu Recht über die außerordentlichen Errungenschaften von Wissenschaft und Technik und sind dankbar dafür, dass dadurch zahllose Übel, die das menschliche Leben heimsuchten und großes Leid verursachten, beseitigt werden konnten. Gleichzeitig legen die wissenschaftlichen und technischen Fortschritte, die eine noch nie dagewesene Kontrolle über die Wirklichkeit ermöglichen, eine Vielzahl von Möglichkeiten in die Hände der Menschen, von denen einige ein Risiko für das Überleben der Menschen und eine Gefahr für das gemeinsame Haus darstellen können[3].

Die bemerkenswerten Fortschritte in den neuen Informationstechnologien, insbesondere im digitalen Bereich, bergen daher erstaunliche Möglichkeiten und ernsthafte Risiken, mit schwerwiegenden Auswirkungen auf das Streben nach Gerechtigkeit und Harmonie zwischen den Völkern. Es müssen daher einige dringende Fragen gestellt werden. Was sind die mittel- und langfristigen Folgen der neuen digitalen Technologien? Und welche Auswirkungen werden sie auf das Leben der Einzelnen und der Gesellschaft, auf die internationale Stabilität und den Frieden haben?

### *2. Die Zukunft der künstlichen Intelligenz zwischen Verheißung und Risiko*

Die Fortschritte in der Informationstechnologie und die Entwicklung digitaler Technologien in den letzten Jahrzehnten haben bereits zu tiefgreifenden Veränderungen in der globalen Gesellschaft und ihrer Dynamik geführt. Neue digitale Instrumente verändern das Gesicht der Kommunikation, der öffentlichen Verwaltung, der Bildung, des Konsums, des persönlichen Austauschs und unzähliger anderer Aspekte des täglichen Lebens.

Darüber hinaus können Technologien, die eine Vielzahl von Algorithmen einsetzen, aus den digitalen Spuren, die im Internet hinterlassen werden, Daten extrahieren, die es ermöglichen, die Denk- und Beziehungsgewohnheiten der Menschen, oft ohne ihr Wissen, zu kommerziellen oder politischen Zwecken zu kontrollieren, wodurch die bewusste Ausübung der Entscheidungsfreiheit eingeschränkt wird. In einem Raum wie dem Internet, der durch eine Informationsflut gekennzeichnet ist, können sie nämlich den Datenfluss nach Auswahlkriterien strukturieren, die der Nutzer nicht immer wahrnimmt.

Wir müssen daran erinnern, dass wissenschaftliche Forschung und technologische Innovationen nicht losgelöst von der Realität und „neutral“[4], sondern kulturellen Einflüssen unterworfen sind. Insofern es sich um ganz und gar menschliche Tätigkeiten handelt, spiegeln die Richtungen, die sie einschlagen, Entscheidungen wider, die durch die persönlichen, sozialen und kulturellen Werte jeder Epoche bedingt sind. Dasselbe gilt für die Ergebnisse, die sie erzielen: Gerade weil sie die Frucht spezifisch menschlicher Zugänge zur sie umgebenden Welt sind, haben sie immer eine ethische Dimension, die eng mit den Entscheidungen derer verbunden sind, die Versuche durchführen und die Produktion auf bestimmte Ziele ausrichten.

Dies gilt auch für die Formen künstlicher Intelligenz. Bis heute gibt es in der Welt der Wissenschaft und Technik keine einheitliche Definition dafür. Der Begriff selbst, der inzwischen in den allgemeinen Sprachgebrauch eingegangen ist, umfasst eine Vielzahl von Wissenschaften, Theorien und Techniken, die darauf abzielen, dass Maschinen in ihrer Funktionsweise die kognitiven Fähigkeiten des Menschen reproduzieren oder imitieren. Die Verwendung des Plurals „Formen der Intelligenz“ kann vor allem dazu beitragen, die unüberbrückbare Kluft zu betonen, die zwischen diesen Systemen, so erstaunlich und leistungsfähig sie auch sein mögen, und dem Menschen besteht: Sie sind letztlich „bruchstückhaft“ in dem Sinne, dass sie nur bestimmte Funktionen der menschlichen Intelligenz imitieren oder reproduzieren können. Die Verwendung des Plurals unterstreicht auch, dass diese untereinander sehr verschiedenen Geräte immer als „soziotechnische Systeme“ betrachtet werden sollten. In der Tat hängt ihre Wirkung – unabhängig von der zugrundeliegenden Technologie – nicht nur davon ab, wie sie konzipiert sind, sondern auch von den Zielen und Interessen derjenigen, die sie besitzen und entwickeln, sowie von den Situationen, in denen sie eingesetzt werden.

Künstliche Intelligenz muss daher als eine Galaxie verschiedener Wirklichkeiten verstanden werden, und wir können nicht *a priori* davon ausgehen, dass ihre Entwicklung einen positiven Beitrag zur Zukunft der Menschheit und zum Frieden zwischen den Völkern leisten wird. Ein solches positives Ergebnis wird nur möglich sein, wenn wir uns als dazu fähig erweisen, verantwortungsbewusst zu handeln und grundlegende menschliche Werte wie »Inklusion, Transparenz, Sicherheit, Gerechtigkeit, Vertraulichkeit und Zuverlässigkeit«[5] zu respektieren.

Es reicht auch nicht aus, bei denjenigen, die Algorithmen und digitale Technologien entwickeln, eine Verpflichtung zu ethischem und verantwortungsvollem Handeln vorauszusetzen. Es müssen Organismen gestärkt oder gegebenenfalls geschaffen werden, die sich mit den neu auftretenden ethischen Fragen befassen und die Rechte derjenigen schützen, die Formen der künstlichen Intelligenz nutzen oder von ihnen beeinflusst werden[6].

Die unermessliche Ausbreitung der Technologie muss daher mit einer angemessenen Heranbildung zur Verantwortung für ihre Entwicklung einhergehen. Freiheit und friedliche Koexistenz sind bedroht, wenn der Mensch der Versuchung von Egoismus, Eigennutz, Profitgier und Machtstreben erliegt. Wir haben daher die Pflicht, unseren Blick zu weiten und die technische und wissenschaftliche Forschung auf das Streben nach Frieden und Gemeinwohl auszurichten, im Dienste der ganzheitlichen Entwicklung des Menschen und der Gemeinschaft[7].

Die einem jeden Menschen innewohnende Würde und die Geschwisterlichkeit, die uns als Glieder der einen Menschheitsfamilie verbindet, müssen die Grundlage für die Entwicklung neuer Technologien bilden und als unbestreitbare Kriterien für deren Bewertung noch vor ihrem Einsatz dienen, damit der digitale Fortschritt unter Wahrung der Gerechtigkeit stattfinden und zur Sache des Friedens beitragen kann. Technologische Entwicklungen, die nicht zu einer Verbesserung der Lebensqualität der gesamten Menschheit führen, sondern im Gegenteil Ungleichheiten und Konflikte verschärfen, können niemals als echter Fortschritt angesehen werden[8].

Künstliche Intelligenz wird zunehmend an Bedeutung gewinnen. Die Herausforderungen, die sie mit sich bringt, sind technischer, aber auch anthropologischer, didaktischer, sozialer und politischer Natur. Sie verspricht zum Beispiel das Ersparen schwerer Arbeit, effizientere Produktion, einfacheren Transport und dynamischere Märkte ebenso wie eine Revolution bei der Datenerfassung, -organisation und -überprüfung. Wir müssen uns der rasanten Veränderungen, die jetzt stattfinden, bewusst sein und sie so steuern, dass die grundlegenden Menschenrechte gewahrt bleiben und die Institutionen und Gesetze, die eine ganzheitliche menschliche

Entwicklung fördern, respektiert werden. Künstliche Intelligenz sollte dem besten menschlichen Potenzial und unseren höchsten Zielen dienen, nicht mit ihnen konkurrieren.

### 3. Die Technologie der Zukunft: Maschinen, die von selbst lernen

Künstliche Intelligenz, die auf maschinellen Lerntechniken basiert, befindet sich zwar noch in der Pionierphase, führt aber bereits in ihren vielfältigen Formen zu bedeutenden Veränderungen im gesellschaftlichen Gefüge und übt einen tiefgreifenden Einfluss auf Kulturen, soziales Verhalten und Friedensstiftung aus.

Entwicklungen wie maschinelles Lernen oder *Deep Learning* werfen Fragen auf, die über den Bereich der Technologie und des Ingenieurwesens hinausgehen und mit einem Verständnis zu tun haben, das eng mit dem Sinn des menschlichen Lebens, den grundlegenden Prozessen des Wissens und der Fähigkeit des Geistes, zur Wahrheit zu gelangen, verbunden ist.

Die Fähigkeit einiger Geräte, syntaktisch und semantisch kohärente Texte zu produzieren, ist zum Beispiel keine Garantie für Zuverlässigkeit. Man sagt ihnen nach, dass sie „halluzinieren“ können, d. h. Aussagen generieren können, die auf den ersten Blick plausibel erscheinen, in Wirklichkeit aber unbegründet sind oder Vorurteile weitertragen. Dies stellt ein ernstes Problem dar, wenn künstliche Intelligenz in Desinformationskampagnen eingesetzt wird, die falsche Nachrichten verbreiten und zu einem wachsenden Misstrauen gegenüber den Medien führen. Vertraulichkeit, Dateneigentum und geistiges Eigentum sind weitere Bereiche, in denen die betreffenden Technologien ernsthafte Risiken bergen, zu denen noch weitere negative Folgen ihres Missbrauchs hinzukommen, wie Diskriminierung, Einmischung in Wahlprozesse, das Aufkommen einer Überwachungsgesellschaft, digitale Ausgrenzung und die Verschärfung eines Individualismus, der sich zunehmend von der Gemeinschaft abkoppelt. All diese Faktoren bergen die Gefahr, Konflikte zu schüren und den Frieden zu behindern.

### 4. Das Gespür für Grenzen im technokratischen Paradigma

Unsere Welt ist zu groß, zu vielfältig und zu komplex, um sie vollständig kennen und klassifizieren zu können. Der menschliche Verstand vermag ihren Reichtum niemals auszuschöpfen, auch nicht mit Hilfe der fortschrittlichsten Algorithmen. Diese bieten nämlich keine gesicherten Vorhersagen für die Zukunft, sondern nur statistische Annäherungen. Nicht alles lässt sich vorhersagen, nicht alles lässt sich berechnen; letztlich steht »die Wirklichkeit [...] über der Idee«[9], und wie großartig unsere Rechenkapazität auch sein mag, es wird immer einen unzugänglichen Rest geben, der sich jedem Versuch der Quantifizierung entzieht.

Außerdem ist die große Menge an Daten, die von künstlichen Intelligenzen analysiert werden, an sich noch keine Garantie für Unparteilichkeit. Wenn Algorithmen Informationen extrapolieren, laufen sie immer Gefahr, diese zu verzerren und die Ungerechtigkeiten und Vorurteile des Umfelds, aus dem sie stammen, zu reproduzieren. Je schneller und komplexer sie werden, desto schwieriger ist es zu verstehen, warum sie ein bestimmtes Ergebnis hervorgebracht haben.

„Intelligente“ Maschinen mögen die ihnen zugewiesenen Aufgaben mit zunehmender Effizienz erfüllen, aber der Zweck und der Sinn ihrer Operationen werden weiterhin von Menschen, die ihr je persönliches Werteuniversum besitzen, bestimmt oder ermöglicht. Es besteht die Gefahr, dass die Kriterien, die bestimmten Entscheidungen zugrunde liegen, unklarer werden, dass die Verantwortung für Entscheidungen verschleiert wird und dass die Produzenten sich ihrer Verpflichtung entziehen, zum Wohle der Gemeinschaft zu handeln. In gewisser Weise wird dies durch das technokratische System begünstigt, das die Wirtschaft mit der Technologie verbindet und das Kriterium der Effizienz begünstigt, indem es dazu neigt, alles zu ignorieren, was nicht mit seinen unmittelbaren Interessen zu tun hat[10].

Dies muss uns dazu veranlassen, über einen Aspekt nachzudenken, der in der heutigen technokratischen und effizienzorientierten Mentalität so oft vernachlässigt wird und dennoch für die persönliche und soziale Entwicklung entscheidend ist: das „Gespür für Grenzen“. Wenn der Mensch, der definitionsgemäß sterblich ist, nämlich meint, mit Hilfe der Technik jede Grenze zu überschreiten, läuft er durch die Besessenheit alles

kontrollieren zu wollen Gefahr, die Kontrolle über sich selbst zu verlieren; auf der Suche nach absoluter Freiheit in die Spirale einer technologischen Diktatur zu geraten. Das Anerkennen und Akzeptieren der eigenen geschöpflichen Grenzen ist für den Menschen die unverzichtbare Bedingung, um die Fülle als Gabe zu erlangen, oder besser, anzunehmen. Stattdessen könnten im ideologischen Kontext eines technokratischen Paradigmas, das von der prometheischen Anmaßung der Autarkie beseelt ist, die Ungleichheiten ins Unermessliche wachsen und sich Wissen und Reichtum in den Händen einiger weniger anhäufen, was ernsthafte Risiken für die demokratischen Gesellschaften und das friedliche Zusammenleben mit sich bringt[11].

### 5. *Brisante Themen für die Ethik*

In Zukunft könnte die Zuverlässigkeit eines Hypothekensuchers, die Eignung einer Person für eine Arbeit, die Wahrscheinlichkeit der Rückfälligkeit eines Verurteilten oder das Recht, politisches Asyl oder Sozialhilfe zu erhalten, von Systemen künstlicher Intelligenz bestimmt werden. Das Fehlen unterschiedlicher Vermittlungsebenen, das diese Systeme mit sich bringen, ist für bestimmte Formen von Vorurteilen und Diskriminierung besonders anfällig: Systemfehler können sich leicht vervielfachen und so nicht nur in Einzelfällen zu Ungerechtigkeiten, sondern durch einen Dominoeffekt auch zu echten Formen sozialer Ungleichheit führen.

Darüber hinaus scheinen Formen künstlicher Intelligenz manchmal in der Lage zu sein, die Entscheidungen der Einzelnen durch vorgegebene Optionen, die mit Anreizen und Abschreckungen verbunden sind, oder durch Systeme zur Lenkung persönlicher Entscheidungen, die auf der Aufbereitung von Informationen beruhen, zu beeinflussen. Diese Formen der Manipulation oder sozialer Kontrolle bedürfen sorgfältiger Aufmerksamkeit und Überwachung und implizieren eine klare rechtliche Verantwortung seitens der Hersteller, der Nutzer und der Regierungsbehörden.

Sich automatisierten Prozessen anzuvertrauen, die Individuen kategorisieren, zum Beispiel durch den allgegenwärtigen Einsatz von Überwachungssystemen oder die Einführung von Systemen zur Ermittlung sozialer Bonität, könnte auch tiefgreifende Auswirkungen auf das zivilgesellschaftliche Gefüge haben, indem unangemessene Rangordnungen unter den Bürgern aufgestellt werden. Und diese künstlichen Ranking-Prozesse könnten auch zu Machtkonflikten führen, da sie nicht nur virtuelle Adressaten betreffen, sondern Menschen aus Fleisch und Blut. Die grundlegende Achtung der Menschenwürde verlangt, die Gleichsetzung der Einzigartigkeit der Person mit einem Datensatz abzulehnen. Algorithmen darf nicht erlaubt werden, die Art und Weise zu bestimmen, wie wir die Menschenrechte verstehen, die Grundwerte des Mitgefühls, der Barmherzigkeit und der Vergebung beiseite zu schieben oder die Möglichkeit auszuschließen, dass ein Individuum sich ändert und die Vergangenheit hinter sich lässt.

In diesem Zusammenhang kommen wir nicht umhin, über die Auswirkungen der neuen Technologien auf das Arbeitsleben nachzudenken: Tätigkeiten, die früher ausschließlich der menschlichen Arbeitskraft vorbehalten waren, werden rasch von industriellen Anwendungen der künstlichen Intelligenz übernommen. Auch in diesem Fall besteht das erhebliche Risiko eines unverhältnismäßigen Vorteils für einige wenige zum Preis der Verarmung vieler. Die Achtung der Würde der Arbeitnehmer und die Bedeutung der Beschäftigung für den wirtschaftlichen Wohlstand der Personen, der Familien und der Gesellschaften, die Sicherheit der Arbeitsplätze und faire Gehälter sollten für die internationale Gemeinschaft eine hohe Priorität darstellen, während diese Formen der Technologie immer tiefer in die Arbeitswelt eindringen.

### 6. *Werden wir Schwerter zu Pflugscharen machen?*

Wenn man heutzutage die Welt um uns herum betrachtet, kann man sich den ernstesten ethischen Fragen im Zusammenhang mit der Rüstungsindustrie nicht entziehen. Die Möglichkeit, militärische Operationen mittels ferngesteuerter Systeme durchzuführen, hat zu einer verringerten Wahrnehmung der von ihnen verursachten Zerstörungen und der Verantwortung für ihren Einsatz geführt, was zu einer noch kälteren und distanzierteren Haltung gegenüber der gewaltigen Tragik des Krieges beiträgt. Die Forschung im Bereich neuer Technologien für die so genannten „tödlichen autonomen Waffensysteme“, einschließlich des Einsatzes von künstlicher Intelligenz im Krieg, ist ein ernstester Grund für ethische Bedenken. Autonome Waffensysteme werden niemals

moralisch verantwortliche Subjekte sein können: Die ausschließlich menschliche Fähigkeit zum moralischen Urteil und zur ethischen Entscheidungsfindung ist mehr als ein komplexer Satz von Algorithmen, und diese Fähigkeit kann nicht auf die Programmierung einer Maschine reduziert werden, die, wie „intelligent“ sie auch sein mag, doch immer eine Maschine bleibt. Aus diesem Grund ist es unerlässlich, eine sachgemäße, maßgebliche und kohärente menschliche Kontrolle der Waffensysteme zu garantieren.

Wir können auch nicht die Möglichkeit vernachlässigen, dass hochentwickelte Waffen in die falschen Hände geraten und zum Beispiel Terroranschläge oder Einsätze zur Destabilisierung rechtmäßiger Regierungsinstitutionen erleichtern. Kurz gesagt, die Welt hat es wirklich nicht nötig, dass die neuen Technologien zu einer unfairen Entwicklung des Waffenmarktes und -handels beitragen und so den Wahnsinn des Krieges fördern. Auf diese Weise läuft nicht nur die Intelligenz des Menschen, sondern auch das Herz selbst Gefahr, immer „künstlicher“ zu werden. Die fortschrittlichsten technischen Anwendungen sind nicht einzusetzen, um gewaltsame Konfliktlösungen zu erleichtern, sondern um die Wege des Friedens zu ebnen.

In einer positiveren Betrachtungsweise könnte künstliche Intelligenz, wenn sie zur Förderung einer ganzheitlichen menschlichen Entwicklung eingesetzt würde, wichtige Innovationen in der Landwirtschaft, der Bildung und der Kultur, eine Verbesserung des Lebensstandards ganzer Nationen und Völker sowie das Wachstum der menschlichen Geschwisterlichkeit und der sozialen Freundschaft bewirken. Letztlich ist die Art und Weise, wie wir sie nutzen, um die Geringsten einzubeziehen, d.h. unsere schwächsten und bedürftigsten Brüder und Schwestern, der Maßstab, der unsere Menschlichkeit aufzeigt.

Eine menschliche Sichtweise und der Wunsch nach einer besseren Zukunft für unsere Welt führen zur Notwendigkeit eines interdisziplinären Dialogs, der auf ein ethisches Vorgehen für die Entwicklung von Algorithmen zielt – die *Algor-Ethik* –, bei der die Werte die Richtung für die neuen Technologien weisen[12]. Ethische Fragen sollten vom Beginn der Forschung an berücksichtigt werden, ebenso in den Phasen des Erprobens, des Entwickelns, der Produktion, der Logistik und der Vermarktung. Dies ist der Ansatz der *Ethics by Design*, bei der den Bildungseinrichtungen und den Verantwortlichen des Entscheidungsprozesses eine wesentliche Rolle zukommt.

### 7. Herausforderungen für die Bildung

Die Entwicklung einer Technologie, die die Menschenwürde respektiert und ihr dient, hat deutliche Auswirkungen auf die Bildungseinrichtungen und die Welt der Kultur. Durch die Vervielfachung der Kommunikationsmöglichkeiten haben die digitalen Technologien neue Formen der Begegnung ermöglicht. Es besteht jedoch die Notwendigkeit, fortlaufend über die Art der Beziehungen nachzudenken, zu denen sie uns führen. Die jungen Menschen wachsen in einem kulturellen Umfeld auf, das von der Technologie durchdrungen ist, was unweigerlich einige Fragen bezüglich der Lehr- und Ausbildungsmethoden aufwirft.

Zu lehren, Formen künstlicher Intelligenz zu nutzen, sollte vor allem darauf abzielen, das kritische Denken zu fördern. Es ist notwendig, dass die Nutzer aller Altersgruppen, vor allem aber junge Menschen, eine Fähigkeit entwickeln, Daten und Inhalte, die im Internet abgerufen wurden oder von Systemen der künstlichen Intelligenz erzeugt worden sind, kritisch zu verwenden. Die Schulen, die Universitäten und die wissenschaftlichen Gemeinschaften sind aufgerufen, den Studenten und Berufstätigen dabei zu helfen, sich die sozialen und ethischen Aspekte der Entwicklung und der Nutzung der Technologie anzueignen.

Dazu auszubilden, die neuen Kommunikationsmittel zu verwenden, sollte nicht nur die Fehlinformation, die *Fake News* berücksichtigen, sondern auch das beunruhigende Zunehmen »angestammte[r] Ängste, [...]. Sie haben sich [...] zu verbergen gewusst und vermochten sich hinter neuen Technologien zu potenzieren«[13]. Leider müssen wir wieder einmal gegen die Versuchung ankämpfen, »eine Kultur der Mauern zu errichten, Mauern hochzuziehen, um [die] Begegnung mit anderen Kulturen, mit anderen Menschen«[14] und die Entwicklung eines friedlichen und geschwisterlichen Zusammenlebens zu verhindern.

### 8. Herausforderungen für die Entwicklung des Völkerrechts

Die globale Reichweite der künstlichen Intelligenz macht deutlich, dass neben der Verantwortung der souveränen Staaten, deren Einsatz innerhalb ihres eigenen Hoheitsgebiets zu regeln, internationale Organisationen eine entscheidende Rolle beim Abschluss multilateraler Vereinbarungen spielen können und dabei, deren Anwendung und Umsetzung zu koordinieren[15]. In dieser Hinsicht fordere ich die Völkergemeinschaft auf, gemeinsam daran zu arbeiten, einen verbindlichen internationalen Vertrag zu schließen, der die Entwicklung und den Einsatz von künstlicher Intelligenz in ihren vielfältigen Formen regelt. Das Ziel der Regulierung sollte natürlich nicht nur die Verhinderung schädlicher Praktiken sein, sondern auch die Ermutigung zu einer guten Praxis, indem neue und kreative Ansätze angeregt sowie persönliche und gemeinschaftliche Initiativen erleichtert werden[16].

Letztlich ist es bei der Suche nach normativen Regelungen, die den Entwicklern digitaler Technologien eine ethische Orientierung bieten können, unerlässlich, die menschlichen Werte zu identifizieren, die den Bemühungen der Gesellschaften zugrunde liegen sollten, um die notwendigen gesetzlichen Rahmenbedingungen zu formulieren, zu beschließen und anzuwenden. Das Erarbeiten ethischer Richtlinien für die Entwicklung künstlicher Intelligenz kann nicht davon absehen, die tieferen Fragen nach dem Sinn der menschlichen Existenz, dem Schutz der grundlegenden Menschenrechte und dem Streben nach Gerechtigkeit und Frieden zu berücksichtigen. Dieser Prozess ethischer und rechtlicher Unterscheidung kann eine wertvolle Gelegenheit bieten, um gemeinsam darüber nachzudenken, welche Rolle die Technologie in unserem individuellen und gemeinschaftlichen Leben spielen sollte und wie ihr Einsatz zur Schaffung einer gerechteren und menschlicheren Welt beitragen kann. Aus diesem Grund sollten die Stimmen aller betroffenen Gruppen in den Debatten über die Regulierung der künstlichen Intelligenz berücksichtigt werden, auch die Armen, die Ausgegrenzten und andere, die in globalen Entscheidungsprozessen oft ungehört bleiben.

\* \* \* \* \*

Ich hoffe, dass diese Überlegungen dazu ermutigen, dafür zu sorgen, dass der Fortschritt bei der Entwicklung von Formen künstlicher Intelligenz letztlich der Sache der menschlichen Geschwisterlichkeit und des Friedens dient. Dies ist nicht die Verantwortung einiger weniger, sondern der gesamten Menschheitsfamilie. Der Friede ist nämlich die Frucht von Beziehungen, die den anderen in seiner unveräußerlichen Würde anerkennen und annehmen, sowie von Zusammenarbeit und Engagement bei der Suche nach der ganzheitlichen Entwicklung aller Menschen und aller Völker.

Mein Gebet zu Beginn des neuen Jahres ist, dass die rapide Entwicklung von Formen künstlicher Intelligenz die vielen Ungleichheiten und Ungerechtigkeiten, die es in der Welt bereits gibt, nicht noch vergrößert, sondern dazu beiträgt, Kriege und Konflikte zu beenden und viele Formen des Leidens zu lindern, die die Menschheitsfamilie heimsuchen. Mögen die Christen, die Gläubigen der verschiedenen Religionen und die Männer und Frauen guten Willens in Harmonie zusammenarbeiten, um die Chancen zu nutzen und sich den durch die digitale Revolution verursachten Herausforderungen zu stellen und um den künftigen Generationen eine solidarischere, gerechtere und friedlichere Welt zu übergeben.

Aus dem Vatikan, am 8. Dezember 2023

FRANZISKUS

---

[1] Nr. 33.

[2] *Ebd.*, Nr. 57.

[3] Vgl. Enzyklika *Laudato si'* (24. Mai 2015), 104.

[4] Vgl. *ebd.*, 114.

[5] *Ansprache an die Teilnehmer der Begegnung der „Minerva Dialogues“* (27. März 2023).

[6] Vgl. *ebd.*

[7] *Botschaft an den Vorstandsvorsitzenden des „World Economic Forum“ in Davos-Klosters* (12. Januar 2018).

[8] Vgl. Enzyklika *Laudato si'*, 194; *Ansprache an die Teilnehmer des Seminars „Das Gemeinwohl im digitalen Zeitalter“* (27. September 2019).

[9] Apostolisches Schreiben *Evangelii gaudium* (24 Novembre 2013), 233.

[10] Vgl. Enzyklika *Laudato si'*, 54.

[11] Vgl. *Ansprache an die Teilnehmer der Vollversammlung der Päpstlichen Akademie für das Leben* (28. Februar 2020).

[12] Vgl. *ebd.*

[13] Enzyklika *Fratelli tutti* (3. Oktober 2020), 27.

[14] Vgl. *ebd.*

[15] Vgl. *ebd.*, 170-175.

[16] Vgl. Enzyklika *Laudato si'*, 177.

[01927-DE.01] [Originalsprache: Italienisch]

### **Traduzione in lingua spagnola**

#### ***Inteligencia artificial y paz***

Al iniciar el año nuevo, tiempo de gracia que el Señor nos da a cada uno de nosotros, quisiera dirigirme al Pueblo de Dios, a las naciones, a los Jefes de Estado y de Gobierno, a los Representantes de las distintas religiones y de la sociedad civil, y a todos los hombres y mujeres de nuestro tiempo para expresarles mis mejores deseos de paz.

#### ***1. El progreso de la ciencia y de la tecnología como camino hacia la paz***

La Sagrada Escritura atestigua que Dios ha dado a los hombres su Espíritu para que tengan «habilidad, talento y experiencia en la ejecución de toda clase de trabajos» (*Ex 35,31*). La inteligencia es expresión de la dignidad que nos ha dado el Creador al hacernos a su imagen y semejanza (cf. *Gn 1,26*) y nos ha hecho capaces de responder a su amor a través de la libertad y del conocimiento. La ciencia y la tecnología manifiestan de modo particular esta cualidad fundamentalmente relacional de la inteligencia humana, ambas son producto extraordinario de su potencial creativo.

En la Constitución pastoral *Gaudium et spes*, el Concilio Vaticano II ha insistido en esta verdad, declarando que «siempre se ha esforzado el hombre con su trabajo y con su ingenio en perfeccionar su vida».[1] Cuando los seres humanos, «con ayuda de los recursos técnicos», se esfuerzan para que la tierra «llegue a ser morada digna de toda la familia humana», [2] actúan según el designio de Dios y cooperan con su voluntad de llevar a cumplimiento la creación y difundir la paz entre los pueblos. Asimismo, el progreso de la ciencia y de la técnica,

en la medida en que contribuye a un mejor orden de la sociedad humana y a acrecentar la libertad y la comunión fraterna, lleva al perfeccionamiento del hombre y a la transformación del mundo.

Nos alegramos justamente y agradecemos las extraordinarias conquistas de la ciencia y de la tecnología, gracias a las cuales se ha podido poner remedio a innumerables males que afectaban a la vida humana y causaban grandes sufrimientos. Al mismo tiempo, los progresos técnico-científicos, haciendo posible el ejercicio de un control sobre la realidad, nunca visto hasta ahora, están poniendo en las manos del hombre una vasta gama de posibilidades, algunas de las cuales representan un riesgo para la supervivencia humana y un peligro para la casa común.[3]

Los notables progresos de las nuevas tecnologías de la información, especialmente en la esfera digital, presentan, por tanto, entusiasmantes oportunidades y graves riesgos, con serias implicaciones para la búsqueda de la justicia y de la armonía entre los pueblos. Por consiguiente, es necesario plantearse algunas preguntas urgentes. ¿Cuáles serán las consecuencias, a medio y a largo plazo, de las nuevas tecnologías digitales? ¿Y qué impacto tendrán sobre la vida de los individuos y de la sociedad, sobre la estabilidad internacional y sobre la paz?

## 2. El futuro de la inteligencia artificial entre promesas y riesgos

Los progresos de la informática y el desarrollo de las tecnologías digitales en los últimos decenios ya han comenzado a producir profundas transformaciones en la sociedad global y en sus dinámicas. Los nuevos instrumentos digitales están cambiando el rostro de las comunicaciones, de la administración pública, de la instrucción, del consumo, de las interacciones personales y de otros innumerables aspectos de la vida cotidiana.

Además, las tecnologías que usan un gran número de algoritmos pueden extraer, de los rastros digitales dejados en internet, datos que permiten controlar los hábitos mentales y relacionales de las personas con fines comerciales o políticos, frecuentemente sin que ellos lo sepan, limitándoles el ejercicio consciente de la libertad de elección. De hecho, en un espacio como la web, caracterizado por una sobrecarga de información, se puede estructurar el flujo de datos según criterios de selección no siempre percibidos por el usuario.

Debemos recordar que la investigación científica y las innovaciones tecnológicas no están desencarnadas de la realidad ni son «neutrales», [4] sino que están sujetas a las influencias culturales. En cuanto actividades plenamente humanas, las direcciones que toman reflejan decisiones condicionadas por los valores personales, sociales y culturales de cada época. Lo mismo se diga de los resultados que consiguen. Estas, precisamente en cuanto fruto de planteamientos específicamente humanos hacia el mundo circunstante, tienen siempre una dimensión ética, estrictamente ligada a las decisiones de quien proyecta la experimentación y enfoca la producción hacia objetivos particulares.

Esto vale también para las formas de inteligencia artificial, para la cual, hasta hoy, no existe una definición unívoca en el mundo de la ciencia y de la tecnología. El término mismo, que ha entrado ya en el lenguaje común, abraza una variedad de ciencias, teorías y técnicas dirigidas a hacer que las máquinas reproduzcan o imiten, en su funcionamiento, las capacidades cognitivas de los seres humanos. Hablar en plural de “formas de inteligencia” puede ayudar a subrayar sobre todo la brecha infranqueable que existe entre estos sistemas y la persona humana, por más sorprendentes y potentes que sean. Estos son, a fin de cuentas, “fragmentarios”, en el sentido de que sólo pueden imitar o reproducir algunas funciones de la inteligencia humana. El uso del plural pone en evidencia además que estos dispositivos, muy distintos entre sí, se deben considerar siempre como “sistemas socio-técnicos”. En efecto, su impacto, independientemente de la tecnología de base, no sólo depende del proyecto, sino también de los objetivos y de los intereses del que los posee y del que los desarrolla, así como de las situaciones en las que se usan.

La inteligencia artificial, por tanto, debe ser entendida como una galaxia de realidades distintas y no podemos presumir *a priori* que su desarrollo aporte una contribución benéfica al futuro de la humanidad y a la paz entre los pueblos. Tal resultado positivo sólo será posible si somos capaces de actuar de forma responsable y de

respetar los valores humanos fundamentales como «la inclusión, la transparencia, la seguridad, la equidad, la privacidad y la responsabilidad».[5]

No basta ni siquiera suponer, de parte de quien proyecta algoritmos y tecnologías digitales, un compromiso de actuar de forma ética y responsable. Es preciso reforzar o, si es necesario, instituir organismos encargados de examinar las cuestiones éticas emergentes y de tutelar los derechos de los que utilizan formas de inteligencia artificial o reciben su influencia.[6]

La inmensa expansión de la tecnología, por consiguiente, debe ser acompañada, para su desarrollo, por una adecuada formación en la responsabilidad. La libertad y la convivencia pacífica están amenazadas cuando los seres humanos ceden a la tentación del egoísmo, del interés personal, del afán de lucro y de la sed de poder. Tenemos por ello el deber de ensanchar la mirada y de orientar la búsqueda técnico-científica hacia la consecución de la paz y del bien común, al servicio del desarrollo integral del hombre y de la comunidad.[7]

La dignidad intrínseca de cada persona y la fraternidad que nos vincula como miembros de una única familia humana, deben estar en la base del desarrollo de las nuevas tecnologías y servir como criterios indiscutibles para valorarlas antes de su uso, de modo que el progreso digital pueda realizarse en el respeto de la justicia y contribuir a la causa de la paz. Los desarrollos tecnológicos que no llevan a una mejora de la calidad de vida de toda la humanidad, sino que, por el contrario, agravan las desigualdades y los conflictos, no podrán ser considerados un verdadero progreso.[8]

La inteligencia artificial será cada vez más importante. Los desafíos que plantea no son sólo técnicos, sino también antropológicos, educativos, sociales y políticos. Promete, por ejemplo, un ahorro de esfuerzos, una producción más eficiente, transportes más ágiles y mercados más dinámicos, además de una revolución en los procesos de recopilación, organización y verificación de los datos. Es necesario ser conscientes de las rápidas transformaciones que están ocurriendo y gestionarlas de modo que se puedan salvaguardar los derechos humanos fundamentales, respetando las instituciones y las leyes que promueven el desarrollo humano integral. La inteligencia artificial debería estar al servicio de un mejor potencial humano y de nuestras más altas aspiraciones, no en competencia con ellos.

### 3. La tecnología del futuro: máquinas que aprenden solas

En sus múltiples formas la inteligencia artificial, basada en técnicas de aprendizaje automático (*machine learning*), aunque se encuentre todavía en una fase pionera, ya está introduciendo cambios notables en el tejido de las sociedades, ejercitando una profunda influencia en las culturas, en los comportamientos sociales y en la construcción de la paz.

Desarrollos como el *machine learning* o como el aprendizaje profundo (*deep learning*) plantean cuestiones que trascienden los ámbitos de la tecnología y de la ingeniería y tienen que ver con una comprensión estrictamente conectada con el significado de la vida humana, los procesos básicos del conocimiento y la capacidad de la mente de alcanzar la verdad.

La habilidad de algunos dispositivos para producir textos sintáctica y semánticamente coherentes, por ejemplo, no es garantía de confiabilidad. Se dice que pueden “alucinar”, es decir, generar afirmaciones que a primera vista parecen plausibles, pero que en realidad son infundadas o delatan prejuicios. Esto crea un serio problema cuando la inteligencia artificial se emplea en campañas de desinformación que difunden noticias falsas y llevan a una creciente desconfianza hacia los medios de comunicación. La confidencialidad, la posesión de datos y la propiedad intelectual son otros ámbitos en los que las tecnologías en cuestión plantean graves riesgos, a los que se añaden ulteriores consecuencias negativas unidas a su uso impropio, como la discriminación, la interferencia en los procesos electorales, la implantación de una sociedad que vigila y controla a las personas, la exclusión digital y la intensificación de un individualismo cada vez más desvinculado de la colectividad. Todos estos factores corren el riesgo de alimentar los conflictos y de obstaculizar la paz.

### 4. El sentido del límite en el paradigma tecnocrático

Nuestro mundo es demasiado vasto, variado y complejo para poder ser completamente conocido y clasificado. La mente humana nunca podrá agotar su riqueza, ni siquiera con la ayuda de los algoritmos más avanzados. Estos, de hecho, no ofrecen previsiones garantizadas del futuro, sino sólo aproximaciones estadísticas. No todo puede ser pronosticado, no todo puede ser calculado; al final «la realidad es superior a la idea»[9] y, por más prodigiosa que pueda ser nuestra capacidad de cálculo, habrá siempre un residuo inaccesible que escapa a cualquier intento de cuantificación.

Además, la gran cantidad de datos analizados por las inteligencias artificiales no es de por sí garantía de imparcialidad. Cuando los algoritmos extrapolan informaciones, siempre corren el riesgo de distorsionarlas, reproduciendo las injusticias y los prejuicios de los ambientes en los que se originan. Cuanto más veloces y complejos se vuelven, más difícil es comprender porqué han generado un determinado resultado.

Las máquinas inteligentes pueden efectuar las tareas que se les asignan cada vez con mayor eficiencia, pero el fin y el significado de sus operaciones continuarán siendo determinadas o habilitadas por seres humanos que tienen un propio universo de valores. El riesgo es que los criterios que están en la base de ciertas decisiones se vuelvan menos transparentes, que la responsabilidad decisional se oculte y que los productores puedan eludir la obligación de actuar por el bien de la comunidad. En cierto sentido, esto es favorecido por el sistema tecnocrático, que alía la economía con la tecnología y privilegia el criterio de la eficiencia, tendiendo a ignorar todo aquello que no está vinculado con sus intereses inmediatos.[10]

Esto debe hacernos reflexionar sobre el “sentido del límite”, un aspecto a menudo descuidado en la mentalidad actual, tecnocrática y eficientista, y sin embargo decisivo para el desarrollo personal y social. El ser humano, en efecto, mortal por definición, pensando en sobrepasar todo límite gracias a la técnica, corre el riesgo, en la obsesión de querer controlarlo todo, de perder el control de sí mismo, y en la búsqueda de una libertad absoluta, de caer en la espiral de una dictadura tecnológica. Reconocer y aceptar el propio límite de criatura es para el hombre condición indispensable para conseguir o, mejor, para acoger la plenitud como un don. En cambio, en el contexto ideológico de un paradigma tecnocrático, animado por una prometeica presunción de autosuficiencia, las desigualdades podrían crecer de forma desmesurada, y el conocimiento y la riqueza acumularse en las manos de unos pocos, con graves riesgos para las sociedades democráticas y la coexistencia pacífica.[11]

##### *5. Temas candentes para la ética*

En el futuro, la fiabilidad de quien pide un préstamo, la idoneidad de un individuo para un trabajo, la posibilidad de reincidencia de un condenado o el derecho a recibir asilo político o asistencia social podrían ser determinados por sistemas de inteligencia artificial. La falta de niveles diversificados de mediación que estos sistemas introducen está particularmente expuesta a formas de prejuicio y discriminación. Los errores sistémicos pueden multiplicarse fácilmente, produciendo no sólo injusticias en casos concretos sino también, por efecto dominó, auténticas formas de desigualdad social.

Además, con frecuencia las formas de inteligencia artificial parecen capaces de influenciar las decisiones de los individuos por medio de opciones predeterminadas asociadas a estímulos y persuasiones, o mediante sistemas de regulación de las elecciones personales basados en la organización de la información. Estas formas de manipulación o de control social requieren una atención y una supervisión precisas, e implican una clara responsabilidad legal por parte de los productores, de quienes las usan y de las autoridades gubernamentales.

La dependencia de procesos automáticos que clasifican a los individuos, por ejemplo, por medio del uso generalizado de la vigilancia o la adopción de sistemas de crédito social, también podría tener repercusiones profundas en el entramado social, estableciendo categorizaciones impropias entre los ciudadanos. Y estos procesos artificiales de clasificación podrían llevar incluso a conflictos de poder, no sólo en lo que respecta a destinatarios virtuales, sino a personas de carne y hueso. El respeto fundamental por la dignidad humana postula rechazar que la singularidad de la persona sea identificada con un conjunto de datos. No debemos permitir que los algoritmos determinen el modo en el que entendemos los derechos humanos, que dejen a un lado los valores esenciales de la compasión, la misericordia y el perdón o que eliminen la posibilidad de que un

individuo cambie y deje atrás el pasado.

En este contexto, no podemos dejar de considerar el impacto de las nuevas tecnologías en el ámbito laboral. Trabajos que en un tiempo eran competencia exclusiva de la mano de obra humana son rápidamente absorbidos por las aplicaciones industriales de la inteligencia artificial. También en este caso se corre el riesgo sustancial de un beneficio desproporcionado para unos pocos a costa del empobrecimiento de muchos. El respeto de la dignidad de los trabajadores y la importancia de la ocupación para el bienestar económico de las personas, las familias y las sociedades, la seguridad de los empleos y la equidad de los salarios deberían constituir una gran prioridad para la comunidad internacional, a medida que estas formas de tecnología se van introduciendo cada vez más en los lugares de trabajo.

#### 6. *¿Transformaremos las espadas en arados?*

En estos días, mirando el mundo que nos rodea, no podemos eludir las graves cuestiones éticas vinculadas al sector de los armamentos. La posibilidad de conducir operaciones militares por medio de sistemas de control remoto ha llevado a una percepción menor de la devastación que estos han causado y de la responsabilidad en su uso, contribuyendo a un acercamiento aún más frío y distante a la inmensa tragedia de la guerra. La búsqueda de las tecnologías emergentes en el sector de los denominados “sistemas de armas autónomos letales”, incluido el uso bélico de la inteligencia artificial, es un gran motivo de preocupación ética. Los sistemas de armas autónomos no podrán ser nunca sujetos moralmente responsables. La exclusiva capacidad humana de juicio moral y de decisión ética es más que un complejo conjunto de algoritmos, y dicha capacidad no puede reducirse a la programación de una máquina que, aun siendo “inteligente”, no deja de ser siempre una máquina. Por este motivo, es imperioso garantizar una supervisión humana adecuada, significativa y coherente de los sistemas de armas.

Tampoco podemos ignorar la posibilidad de que armas sofisticadas terminen en las manos equivocadas facilitando, por ejemplo, ataques terroristas o acciones dirigidas a desestabilizar instituciones de gobierno legítimas. En resumen, realmente lo último que el mundo necesita es que las nuevas tecnologías contribuyan al injusto desarrollo del mercado y del comercio de las armas, promoviendo la locura de la guerra. Si lo hace así, no sólo la inteligencia, sino el mismo corazón del hombre correrá el riesgo de volverse cada vez más “artificial”. Las aplicaciones técnicas más avanzadas no deben usarse para facilitar la resolución violenta de los conflictos, sino para pavimentar los caminos de la paz.

En una óptica más positiva, si la inteligencia artificial fuese utilizada para promover el desarrollo humano integral, podría introducir importantes innovaciones en la agricultura, la educación y la cultura, un mejoramiento del nivel de vida de enteras naciones y pueblos, el crecimiento de la fraternidad humana y de la amistad social. En definitiva, el modo en que la usamos para incluir a los últimos, es decir, a los hermanos y las hermanas más débiles y necesitados, es la medida que revela nuestra humanidad.

Una mirada humana y el deseo de un futuro mejor para nuestro mundo llevan a la necesidad de un diálogo interdisciplinar destinado a un desarrollo ético de los algoritmos —*la algorética*—, en el que los valores orienten los itinerarios de las nuevas tecnologías.[12] Las cuestiones éticas deberían ser tenidas en cuenta desde el inicio de la investigación, así como en las fases de experimentación, planificación, distribución y comercialización. Este es el enfoque de la ética de la planificación, en el que las instituciones educativas y los responsables del proceso decisional tienen un rol esencial que desempeñar.

#### 7. *Desafíos para la educación*

El desarrollo de una tecnología que respete y esté al servicio de la dignidad humana tiene claras implicaciones para las instituciones educativas y para el mundo de la cultura. Al multiplicar las posibilidades de comunicación, las tecnologías digitales nos han permitido nuevas formas de encuentro. Sin embargo, continúa siendo necesaria una reflexión permanente sobre el tipo de relaciones al que nos está llevando. Los jóvenes están creciendo en ambientes culturales impregnados de la tecnología y esto no puede dejar de cuestionar los métodos de enseñanza y formación.

La educación en el uso de formas de inteligencia artificial debería centrarse sobre todo en promover el pensamiento crítico. Es necesario que los usuarios de todas las edades, pero sobre todo los jóvenes, desarrollen una capacidad de discernimiento en el uso de datos y de contenidos obtenidos en la web o producidos por sistemas de inteligencia artificial. Las escuelas, las universidades y las sociedades científicas están llamadas a ayudar a los estudiantes y a los profesionales a hacer propios los aspectos sociales y éticos del desarrollo y el uso de la tecnología.

La formación en el uso de nuevos instrumentos de comunicación debería considerar no sólo la desinformación, las falsas noticias, sino también el inquietante aumento de «miedos ancestrales que [...] han sabido esconderse y potenciarse detrás de nuevas tecnologías».[13] Lamentablemente, una vez más nos encontramos teniendo que combatir “la tentación de hacer una cultura de muros, de levantar muros para impedir el encuentro con otras culturas, con otra gente”[14] y el desarrollo de una coexistencia pacífica y fraterna.

#### *8. Desafíos para el desarrollo del derecho internacional*

El alcance global de la inteligencia artificial hace evidente que, junto a la responsabilidad de los estados soberanos de disciplinar internamente su uso, las organizaciones internacionales pueden desempeñar un rol decisivo en la consecución de acuerdos multilaterales y en la coordinación de su aplicación y actuación.[15] A este propósito, exhorto a la comunidad de las naciones a trabajar unida para adoptar un tratado internacional vinculante, que regule el desarrollo y el uso de la inteligencia artificial en sus múltiples formas. Naturalmente, el objetivo de la reglamentación no debería ser sólo la prevención de las malas prácticas, sino también alentar las mejores prácticas, estimulando planteamientos nuevos y creativos y facilitando iniciativas personales y colectivas.[16]

En definitiva, en la búsqueda de modelos normativos que puedan proporcionar una guía ética a quienes desarrollan tecnologías digitales, es indispensable identificar los valores humanos que deberían estar en la base del compromiso de las sociedades para formular, adoptar y aplicar los marcos legislativos necesarios. El trabajo de redacción de las orientaciones éticas para la producción de formas de inteligencia artificial no puede prescindir de la consideración de cuestiones más profundas, relacionadas con el significado de la existencia humana, la tutela de los derechos humanos fundamentales y la búsqueda de la justicia y de la paz. Este proceso de discernimiento ético y jurídico puede revelarse como una valiosa ocasión para una reflexión compartida sobre el rol que la tecnología debería tener en nuestra vida personal y comunitaria y sobre cómo su uso podría contribuir a la creación de un mundo más justo y humano. Por este motivo, en los debates sobre la reglamentación de la inteligencia artificial, se debería tener en cuenta la voz de todas las partes interesadas, incluidos los pobres, los marginados y otros más que a menudo quedan sin ser escuchados en los procesos decisionales globales.

\* \* \* \* \*

Espero que esta reflexión anime a hacer que los progresos en el desarrollo de formas de inteligencia artificial contribuyan, en última instancia, a la causa de la fraternidad humana y de la paz. No es responsabilidad de unos pocos, sino de toda la familia humana. La paz, en efecto, es el fruto de relaciones que reconocen y acogen al otro en su dignidad inalienable, y de cooperación y esfuerzo en la búsqueda del desarrollo integral de todas las personas y de todos los pueblos.

Mi oración al comienzo del nuevo año es que el rápido desarrollo de formas de inteligencia artificial no aumente las ya numerosas desigualdades e injusticias presentes en el mundo, sino que ayude a poner fin a las guerras y los conflictos, y a aliviar tantas formas de sufrimiento que afectan a la familia humana. Que los fieles cristianos, los creyentes de distintas religiones y los hombres y mujeres de buena voluntad puedan colaborar en armonía para aprovechar las oportunidades y afrontar los desafíos que plantea la revolución digital, y dejar a las generaciones futuras un mundo más solidario, justo y pacífico.

---

[1] N. 33.

[2] *Ibíd.*, n. 57.

[3] Cf. Carta enc. *Laudato si'* (24 mayo 2015), 104.

[4] Cf. *ibíd.*, 114.

[5] *Discurso a los participantes en el encuentro "Minerva Dialogues"* (27 marzo 2023).

[6] Cf. *ibíd.*

[7] Cf. *Mensaje al Presidente Ejecutivo del "World Economic Forum" en Davos-Klosters* (12 enero 2018).

[8] Cf. Carta enc. *Laudato si'*, 194; *Discurso a los participantes en un Seminario sobre "El bien común en la era digital"* (27 septiembre 2019).

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

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

[11] Cf. *Discurso a los participantes en la Plenaria de la Pontificia Academia para la Vida* (28 febrero 2020).

[12] Cf. *ibíd.*

[13] Carta enc. *Fratelli tutti* (3 octubre 2020), 27.

[14] Cf. *ibíd.*

[15] Cf. *ibíd.*, 170-175.

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

[01927-ES.01] [Texto original: Italiano]

### **Traduzione in lingua portoghese**

#### ***Inteligência artificial e paz***

No início do novo ano, tempo de graça concedido pelo Senhor a cada um de nós, quero dirigir-me ao Povo de Deus, às nações, aos Chefes de Estado e de Governo, aos Representantes das diversas religiões e da sociedade civil, a todos os homens e mulheres do nosso tempo para lhes expressar os meus votos de paz.

#### ***1. O progresso da ciência e da tecnologia como caminho para a paz***

A Sagrada Escritura atesta que Deus deu aos homens o seu Espírito a fim de terem «sabedoria, inteligência e

capacidade para toda a espécie de trabalho» (Ex 35, 31). A inteligência é expressão da dignidade que nos foi dada pelo Criador, que nos fez à sua imagem e semelhança (cf. Gn 1, 26) e nos tornou capazes, através da liberdade e do conhecimento, de responder ao seu amor. Esta qualidade fundamentalmente relacional da inteligência humana manifesta-se de modo particular na ciência e na tecnologia, que são produtos extraordinários do seu potencial criativo.

Na Constituição pastoral *Gaudium et spes*, o Concílio Vaticano II reafirmou esta verdade, declarando que «sempre o homem procurou, com o seu trabalho e engenho, desenvolver mais a própria vida».[1] Quando os seres humanos, «recorrendo à técnica», se esforçam por que a terra «se torne habitação digna para toda a humanidade»,[2] agem segundo o desígnio divino e cooperam com a vontade que Deus tem de levar à perfeição a criação e difundir a paz entre os povos. Assim o próprio progresso da ciência e da técnica – na medida em que contribui para uma melhor organização da sociedade humana, para o aumento da liberdade e da comunhão fraterna – leva ao aperfeiçoamento do homem e à transformação do mundo.

Justamente nos alegramos e sentimos reconhecidos pelas extraordinárias conquistas da ciência e da tecnologia, graças às quais se pôs remédio a inúmeros males que afligiam a vida humana e causavam grandes sofrimentos. Ao mesmo tempo, os progressos técnico-científicos, que permitem exercer um controle – até agora inédito – sobre a realidade, colocam nas mãos do homem um vasto leque de possibilidades, algumas das quais podem constituir um risco para a sobrevivência humana e um perigo para a casa comum.[3]

Deste modo os progressos notáveis das novas tecnologias da informação, sobretudo na esfera digital, apresentam oportunidades entusiasmantes mas também graves riscos, com sérias implicações na prossecução da justiça e da harmonia entre os povos. Por isso torna-se necessário interrogar-nos sobre algumas questões urgentes: quais serão as consequências, a médio e longo prazo, das novas tecnologias digitais? E que impacto terão elas sobre a vida dos indivíduos e da sociedade, sobre a estabilidade e a paz?

## 2. O futuro da inteligência artificial, por entre promessas e riscos

Os progressos da informática e o desenvolvimento das tecnologias digitais, nas últimas décadas, começaram já a produzir profundas transformações na sociedade global e nas suas dinâmicas. Os novos instrumentos digitais estão a mudar a fisionomia das comunicações, da administração pública, da instrução, do consumo, dos intercâmbios pessoais e de inúmeros outros aspetos da vida diária.

Além disso as tecnologias que se servem duma multiplicidade de algoritmos podem, dos vestígios digitais deixados na *internet*, extrair dados que permitem controlar os hábitos mentais e relacionais das pessoas para fins comerciais ou políticos, muitas vezes sem o seu conhecimento, limitando o exercício consciente da sua liberdade de escolha. De facto, num espaço como a *web* caracterizado por uma sobrecarga de informações, pode-se compor o fluxo de dados segundo critérios de seleção nem sempre enxergados pelo utente.

Devemos recordar-nos de que a pesquisa científica e as inovações tecnológicas não estão desencarnadas da realidade nem são «neutras»,[4] mas estão sujeitas às influências culturais. Sendo atividades plenamente humanas, os rumos que tomam refletem opções condicionadas pelos valores pessoais, sociais e culturais de cada época. E o mesmo se diga dos resultados que alcançam: enquanto fruto de abordagens especificamente humanas do mundo envolvente, têm sempre uma dimensão ética, intimamente ligada às decisões de quem projeta a experimentação e orienta a produção para objetivos particulares.

Isto aplica-se também às formas de inteligência artificial. Desta, até ao momento, não existe uma definição unívoca no mundo da ciência e da tecnologia. A própria designação, que já entrou na linguagem comum, abrange uma variedade de ciências, teorias e técnicas destinadas a fazer com que as máquinas, no seu funcionamento, reproduzam ou imitem as capacidades cognitivas dos seres humanos. Falar de «formas de inteligência», no plural, pode ajudar sobretudo a assinalar o fosso intransponível existente entre estes sistemas, por mais surpreendentes e poderosos que sejam, e a pessoa humana: em última análise, aqueles são «fragmentários» já que têm possibilidades de imitar ou reproduzir apenas algumas funções da inteligência humana. Além disso o uso do plural destaca que tais dispositivos, muito diferentes entre si, devem ser sempre

considerados como «sistemas sociotécnicos». Com efeito o seu impacto, independentemente da tecnologia de base, depende não só da projeção, mas também dos objetivos e interesses de quem os possui e de quem os desenvolve, bem como das situações em que são utilizados.

Por conseguinte a inteligência artificial deve ser entendida como uma galáxia de realidades diversas e não podemos presumir a priori que o seu desenvolvimento traga um contributo benéfico para o futuro da humanidade e para a paz entre os povos. O resultado positivo só será possível se nos demonstrarmos capazes de agir de maneira responsável e respeitar valores humanos fundamentais como «a inclusão, a transparência, a segurança, a equidade, a privacidade e a fiabilidade».[5]

E não é suficiente presumir, por parte de quem projeta algoritmos e tecnologias digitais, um empenho por agir de modo ético e responsável. É preciso reforçar ou, se necessário, instituir organismos encarregados de examinar as questões éticas emergentes e tutelar os direitos de quantos utilizam formas de inteligência artificial ou são influenciados por ela.[6]

Assim, a imensa expansão da tecnologia deve ser acompanhada por uma adequada formação da responsabilidade pelo seu desenvolvimento. A liberdade e a convivência pacífica ficam ameaçadas, quando os seres humanos cedem à tentação do egoísmo, do interesse próprio, da ânsia de lucro e da sede de poder. Por isso temos o dever de alargar o olhar e orientar a pesquisa técnico-científica para a prossecução da paz e do bem comum, ao serviço do desenvolvimento integral do homem e da comunidade.[7]

A dignidade intrínseca de cada pessoa e a fraternidade que nos une como membros da única família humana devem estar na base do desenvolvimento de novas tecnologias e servir como critérios indiscutíveis para as avaliar antes da sua utilização, para que o progresso digital possa verificar-se no respeito pela justiça e contribuir para a causa da paz. Os avanços tecnológicos que não conduzem a uma melhoria da qualidade de vida da humanidade inteira, antes pelo contrário agravam as desigualdades e os conflitos, nunca poderão ser considerados um verdadeiro progresso.[8]

A inteligência artificial tornar-se-á cada vez mais importante. Os desafios que coloca não são apenas de ordem técnica, mas também antropológica, educacional, social e política. Deixa esperar, por exemplo, poupança de esforços, produção mais eficiente, transportes mais fáceis e mercados mais dinâmicos, bem como uma revolução nos processos de recolha, organização e verificação de dados. Precisamos de estar conscientes das rápidas transformações em curso e geri-las de forma a salvaguardar os direitos humanos fundamentais, respeitando as instituições e as leis que promovem o progresso humano integral. A inteligência artificial deveria estar ao serviço dum melhor potencial humano e das nossas mais altas aspirações, e não em competição com eles.

### 3. A tecnologia do futuro: máquinas que aprendem sozinhas

Nas suas múltiplas formas, a inteligência artificial, baseada em técnicas de aprendizagem automática (*machine learning*), embora ainda numa fase pioneira, já está a introduzir mudanças notáveis no tecido das sociedades, exercendo uma influência profunda nas culturas, nos comportamentos sociais e na construção da paz.

Desenvolvimentos como a aprendizagem automática (*machine learning*) ou a aprendizagem profunda (*deep learning*) levantam questões que transcendem os âmbitos da tecnologia e da engenharia e têm a ver com uma compreensão intimamente ligada ao significado da vida humana, aos processos basilares do conhecimento e à capacidade que tem a mente de alcançar a verdade.

A capacidade de alguns dispositivos produzirem textos sintática e semanticamente coerentes, por exemplo, não é garantia de fiabilidade. Diz-se que podem «alucinar», isto é, gerar afirmações que à primeira vista parecem plausíveis, mas na realidade são infundadas ou preconceituosas. Isto coloca um sério problema quando a inteligência artificial é utilizada em campanhas de desinformação que espalham notícias falsas e levam a uma desconfiança crescente relativamente aos meios de comunicação. A confidencialidade, a posse dos dados e a propriedade intelectual são outros âmbitos em que as tecnologias em questão comportam graves riscos, aos

quais se vêm juntar outras consequências negativas ligadas a um uso indevido, como a discriminação, a interferência nos processos eleitorais, a formação duma sociedade que vigia e controla as pessoas, a exclusão digital e a exacerbação dum individualismo cada vez mais desligado da coletividade. Todos estes fatores correm o risco de alimentar os conflitos e obstaculizar a paz.

#### 4. *O sentido do limite, no paradigma tecnocrático*

O nosso mundo é demasiado vasto, variado e complexo para ser completamente conhecido e classificado. A mente humana nunca poderá esgotar a sua riqueza, nem sequer com a ajuda dos algoritmos mais avançados. De facto, estes não oferecem previsões garantidas do futuro, mas apenas aproximações estatísticas. Nem tudo pode ser previsto, nem tudo pode ser calculado; no fim de contas, «a realidade é superior à ideia»[9] e, por mais prodigiosa que seja a nossa capacidade de calcular, haverá sempre um resíduo inacessível que escapa a qualquer tentativa de quantificação.

Além disso, a grande quantidade de dados analisados pelas inteligências artificiais não é, por si só, garantia de imparcialidade. Quando os algoritmos extrapolam informações, correm sempre o risco de as distorcer, replicando as injustiças e os preconceitos dos ambientes onde têm origem. Quanto mais rápidos e complexos eles se tornam, mais difícil é compreender por que produziram um determinado resultado.

As máquinas inteligentes podem desempenhar as tarefas que lhes são atribuídas com uma eficiência cada vez maior, mas a finalidade e o significado das suas operações continuarão a ser determinados ou capacitados por seres humanos com o seu próprio universo de valores. O risco é que os critérios subjacentes a certas escolhas se tornem menos claros, que a responsabilidade de decisão seja ocultada e que os produtores possam subtrair-se à obrigação de agir para o bem da comunidade. Em certo sentido, isto é favorecido pelo sistema tecnocrático, que alia a economia à tecnologia e privilegia o critério da eficiência, tendendo a ignorar tudo o que não esteja ligado aos seus interesses imediatos.[10]

Isto deve fazer-nos refletir sobre um aspeto transcurado frequentemente na atual mentalidade tecnocrática e eficientista, mas decisivo para o desenvolvimento pessoal e social: o «sentido do limite». Com efeito o ser humano, mortal por definição, pensando em ultrapassar todo o limite mediante a técnica, corre o risco, na obsessão de querer controlar tudo, de perder o controle sobre si mesmo; na busca duma liberdade absoluta, de cair na espiral duma ditadura tecnológica. Reconhecer e aceitar o próprio limite de criatura é condição indispensável para que o homem alcance ou, melhor, acolha a plenitude como uma dádiva; ao passo que, no contexto ideológico dum paradigma tecnocrático animado por uma prometeica presunção de autossuficiência, as desigualdades poderiam crescer sem medida, e o conhecimento e a riqueza acumular-se nas mãos de poucos, com graves riscos para as sociedades democráticas e uma coexistência pacífica.[11]

#### 5. *Temas quentes para a ética*

No futuro, a fiabilidade de quem solicita um mútuo, a idoneidade dum indivíduo para determinado emprego, a possibilidade de reincidência dum condenado ou o direito a receber asilo político ou assistência social poderão ser determinados por sistemas de inteligência artificial. A falta de níveis diversificados de mediação que tais sistemas introduzem está particularmente exposta a formas de preconceito e discriminação: os erros do sistema podem multiplicar-se facilmente, gerando não só injustiças em casos individuais, mas também, por efeito dominó, verdadeiras formas de desigualdade social.

Além disso, por vezes, as formas de inteligência artificial parecem capazes de influenciar as decisões dos indivíduos através de opções predeterminadas associadas a estímulos e dissuasões, ou então através de sistemas de regulação das opções pessoais baseados na organização das informações. Estas formas de manipulação ou controle social requerem atenção e vigilância cuidadosas, implicando uma clara responsabilidade legal por parte dos produtores, de quem os contrata e das autoridades governamentais.

O ato de se confiar a processos automáticos que dispõem os indivíduos por categorias, por exemplo, através dum uso invasivo da vigilância ou da adoção de sistemas de crédito social, poderia ter repercussões profundas

também no tecido civil, estabelecendo classificações inadequadas entre os cidadãos. E estes processos artificiais de classificação poderiam levar também a conflitos de poder, envolvendo não apenas destinatários virtuais, mas também pessoas de carne e osso. O respeito fundamental pela dignidade humana requer a rejeição de que a unicidade da pessoa seja identificada com um conjunto de dados. Não se deve permitir que os algoritmos determinem o modo como entendemos os direitos humanos, ponham de lado os valores essenciais da compaixão, da misericórdia e do perdão, ou eliminem a possibilidade de um indivíduo mudar e deixar para trás o passado.

Neste contexto, não podemos deixar de considerar o impacto das novas tecnologias no âmbito laboral: trabalhos, que outrora eram prerrogativa exclusiva da mão-de-obra humana, acabam rapidamente absorvidos pelas aplicações industriais da inteligência artificial. Também neste caso, há substancialmente o risco duma vantagem desproporcionada para poucos à custa do empobrecimento de muitos. A Comunidade Internacional, ao ver como tais formas de tecnologia penetram cada vez mais profundamente nos locais de trabalho, deveria considerar como alta prioridade o respeito pela dignidade dos trabalhadores e a importância do emprego para o bem-estar económico das pessoas, das famílias e das sociedades, a estabilidade dos empregos e a equidade dos salários.

## 6. *Transformaremos as espadas em relhas de arado?*

Nestes dias, contemplando o mundo que nos rodeia, não se pode ignorar as graves questões éticas relacionadas com o setor dos armamentos. A possibilidade de efetuar operações militares através de sistemas de controle remoto levou a uma perceção menor da devastação por eles causada e da responsabilidade da sua utilização, contribuindo para uma abordagem ainda mais fria e destacada da imensa tragédia da guerra. A pesquisa sobre as tecnologias emergentes no setor dos chamados «sistemas de armas letais autónomas», incluindo a utilização bélica da inteligência artificial, é um grave motivo de preocupação ética. Os sistemas de armas autónomos nunca poderão ser sujeitos moralmente responsáveis: a exclusiva capacidade humana de julgamento moral e de decisão ética é mais do que um conjunto complexo de algoritmos, e tal capacidade não pode ser reduzida à programação duma máquina que, por mais «inteligente» que seja, permanece sempre uma máquina. Por esta razão, é imperioso garantir uma supervisão humana adequada, significativa e coerente dos sistemas de armas.

Também não podemos ignorar a possibilidade de armas sofisticadas caírem em mãos erradas, facilitando, por exemplo, ataques terroristas ou intervenções visando desestabilizar instituições legítimas de Governo. Em resumo, o mundo não precisa realmente que as novas tecnologias contribuam para o iníquo desenvolvimento do mercado e do comércio das armas, promovendo a loucura da guerra. Ao fazê-lo, não só a inteligência, mas também o próprio coração do homem, correrá o risco de se tornar cada vez mais «artificial». As aplicações técnicas mais avançadas não devem ser utilizadas para facilitar a resolução violenta dos conflitos, mas para pavimentar os caminhos da paz.

Numa ótica mais positiva, se a inteligência artificial fosse utilizada para promover o desenvolvimento humano integral, poderia introduzir inovações importantes na agricultura, na instrução e na cultura, uma melhoria do nível de vida de inteiras nações e povos, o crescimento da fraternidade humana e da amizade social. Em última análise, a forma como a utilizamos para incluir os últimos, isto é, os irmãos e irmãs mais frágeis e necessitados, é a medida reveladora da nossa humanidade.

Um olhar humano e o desejo dum futuro melhor para o nosso mundo levam à necessidade dum diálogo interdisciplinar voltado para um desenvolvimento ético dos algoritmos – a *algor-ética* -, em que sejam os valores a orientar os percursos das novas tecnologias.[12] As questões éticas deveriam ser tidas em consideração desde o início da pesquisa, bem como nas fases de experimentação, projeção, produção, distribuição e comercialização. Esta é a abordagem da ética da projeção, na qual as instituições educativas e os responsáveis pelo processo de decisão têm um papel essencial a desempenhar.

## 7. *Desafios para a educação*

O desenvolvimento duma tecnologia que respeite e sirva a dignidade humana tem implicações claras para as instituições educativas e para o mundo da cultura. Ao multiplicar as possibilidades de comunicação, as tecnologias digitais permitiram encontrar-se de novas formas. Todavia continua a ser necessária uma reflexão contínua sobre o tipo de relações para onde nos estão encaminhando. Os jovens estão a crescer em ambientes culturais impregnados de tecnologia, o que não pode deixar de pôr em causa os métodos de ensino e formação.

A educação para o uso de formas de inteligência artificial deveria visar sobretudo a promoção do pensamento crítico. É necessário que os utentes das várias idades, mas principalmente os jovens, desenvolvam uma capacidade de discernimento no uso de dados e conteúdos recolhidos na *web* ou produzidos por sistemas de inteligência artificial. As escolas, as universidades e as sociedades científicas são chamadas a ajudar os estudantes e profissionais a assumir os aspetos sociais e éticos do progresso e da utilização da tecnologia.

A formação no uso dos novos instrumentos de comunicação deveria ter em conta não só a desinformação, as notícias falsas, mas também a recrudescência preocupante de «medos ancestrais (...) que souberam esconder-se e revigorar-se por detrás das novas tecnologias».[13] Infelizmente, encontramos-nos mais uma vez a combater «a tentação de fazer uma cultura dos muros, de erguer os muros (...), para impedir este encontro com outras culturas, com outras pessoas»[14] e o desenvolvimento duma coexistência pacífica e fraterna.

#### 8. *Desafios para o desenvolvimento do direito internacional*

O alcance global da inteligência artificial deixa claro que, juntamente com a responsabilidade dos Estados soberanos de regular a sua utilização internamente, as Organizações Internacionais podem desempenhar um papel decisivo na obtenção de acordos multilaterais e na coordenação da sua aplicação e implementação.[15] A este respeito, exorto a Comunidade das Nações a trabalhar unida para adotar um tratado internacional vinculativo, que regule o desenvolvimento e o uso da inteligência artificial nas suas variadas formas. Naturalmente o objetivo da regulamentação não deveria ser apenas a prevenção de más aplicações, mas também o incentivo às boas aplicações, estimulando abordagens novas e criativas e facilitando iniciativas pessoais e coletivas.[16]

Em última análise, na busca de modelos normativos que possam fornecer uma orientação ética aos criadores de tecnologias digitais, é indispensável identificar os valores humanos que deveriam estar na base dos esforços das sociedades para formular, adotar e aplicar os quadros legislativos necessários. O trabalho de elaboração de diretrizes éticas para a produção de formas de inteligência artificial não pode prescindir da consideração de questões mais profundas relativas ao significado da existência humana, à proteção dos direitos humanos fundamentais, à busca da justiça e da paz. Este processo de discernimento ético e jurídico pode revelar-se preciosa ocasião para uma reflexão compartilhada sobre o papel que a tecnologia deveria ter na nossa vida individual e comunitária e sobre a forma como a sua utilização possa contribuir para a criação dum mundo mais equitativo e humano. Por este motivo, nos debates sobre a regulamentação da inteligência artificial, dever-se-ia ter em conta as vozes de todas as partes interessadas, incluindo os pobres, os marginalizados e outros que muitas vezes permanecem ignorados nos processos de decisão globais.

\* \* \* \* \*

Espero que esta reflexão encoraje a fazer com que os progressos no desenvolvimento de formas de inteligência artificial sirvam, em última análise, a causa da fraternidade humana e da paz. Não é responsabilidade de poucos, mas da família humana inteira. De facto, a paz é fruto de relações que reconhecem e acolhem o outro na sua dignidade inalienável, e de cooperação e compromisso na busca do desenvolvimento integral de todas as pessoas e de todos os povos.

No início do novo ano, a minha oração é que o rápido desenvolvimento de formas de inteligência artificial não aumente as já demasiadas desigualdades e injustiças presentes no mundo, mas contribua para pôr fim às guerras e conflitos e para aliviar muitas formas de sofrimento que afligem a família humana. Possam os fiéis cristãos, os crentes das várias religiões e os homens e mulheres de boa vontade colaborar harmoniosamente

para aproveitar as oportunidades e enfrentar os desafios colocados pela revolução digital, e entregar às gerações futuras um mundo mais solidário, justo e pacífico.

Vaticano, 8 de dezembro de 2023.

FRANCISCO

---

[1] N. 33.

[2] *Ibid.*, 57.

[3] Cf. Francisco, Carta enc. *Laudato si'* (24/VI/2015), 104.

[4] Cf. *ibid.*, 114.

[5] Francisco, *Discurso aos participantes no Encontro dos «Minerva Dialogues»* (27/III/2023).

[6] Cf. *ibid.*

[7] Cf. Francisco, *Mensagem ao Presidente Executivo do «World Economic Forum» em Davos-Klosters* (12/I/2018).

[8] Cf. Carta enc. *Laudato si'*, 194; Francisco, *Discurso aos participantes no Seminário «O bem comum na era digital»* (27/IX/2019).

[9] Francisco, Exort. ap. *Evangeli gaudium* (24/XI/2013), 233.

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

[11] Cf. Francisco, *Discurso aos participantes na Plenária da Pontifícia Academia em prol da Vida* (28/II/2020).

[12] Cf. *ibid.*

[13] Francisco, Carta enc. *Fratelli tutti* (03/IX/2020), 27.

[14] Cf. *ibid.*

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

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

[01927-PO.01] [Texto original: Italiano]

**Traduzione in lingua polacca**

***Sztuczna inteligencja i pokój***

Na początku Nowego Roku, czasu łaski, który Pan daje każdemu z nas, chciałbym zwrócić się do Ludu Bożego,

narodów, głów państw i rządów, przedstawiciele różnych religii i społeczeństwa obywatelskiego oraz wszystkich mężczyzn i kobiet naszych czasów, aby złożyć najlepsze życzenia pokoju.

### 1. *Postęp nauki i technologii jako droga do pokoju*

Pismo Święte zaświadcza, że Bóg dał ludziom swego Ducha, aby posiadali „mądrość, rozum, wiedzę i znajomość wszelkiego rzemiosła” (*Wj 35, 31*). Rozum jest wyrazem godności nadanej nam przez Stwórcę, który stworzył nas na swój obraz i podobieństwo (por. *Rdz 1, 26*) i umożliwił nam odpowiadanie na Jego miłość poprzez wolność i poznanie. Nauka i technologia w szczególny sposób ukazują tę fundamentalnie relacyjną właściwość ludzkiego rozumu: są one niezwykle wytworami jego potencjału twórczego.

W Konstytucji duszpasterskiej *Gaudium et spes*, Sobór Watykański II potwierdził tę prawdę, oświadczając, że „człowiek zawsze starał się dynamiczniej rozwijać swoje życie dzięki własnej pracy i uzdolnieniom duchowym”[1]. Kiedy ludzie, „za pomocą środków technicznych”, dążą do tego, aby ziemia „stawała się mieszkaniem godnym całej rodziny ludzkiej”[2], działają zgodnie z planem Boga i współpracują z Jego wolą, aby dokończyć dzieło stworzenia i szerzyć pokój między narodami. Także postęp nauki i techniki, o ile przyczynia się do lepszego porządku społeczności ludzkiej, do rozwoju wolności i braterskiej komunii, prowadzi w ten sposób do udoskonalenia człowieka i przemiany świata.

Słusznie cieszymy się i jesteśmy wdzięczni za niezwykle zdobycze nauki i techniki, dzięki którym zażegnano niezliczone bolączki, które trapiły ludzkie życie i powodowały wielkie cierpienia. Jednocześnie postęp technologiczno-naukowy, umożliwiając sprawowanie niespotykanej dotąd kontroli nad rzeczywistością, oddaje w ludzkie ręce szeroki wachlarz możliwości, z których pewne mogą stanowić zagrożenie dla ludzkiego przetrwania i niebezpieczeństwo dla wspólnego domu[3].

Znaczący postęp nowych technologii informacyjnych, zwłaszcza w sferze cyfrowej, stwarza zatem fascynujące szanse i poważne zagrożenia, z ogromnymi konsekwencjami dla dążenia do sprawiedliwości i zgody między ludźmi. Należy zatem zadać kilka pilnych pytań. Jakie będą średnio- i długoterminowe konsekwencje nowych technologii cyfrowych? I jaki wpływ będą one miały na życie jednostek i społeczeństwa, na międzynarodową stabilność i pokój?

### 2. *Przyszłość sztucznej inteligencji między obietnicami a zagrożeniami*

Postęp w informatyce i rozwój technologii cyfrowych, w minionych dekadach, już zaczęły powodować głębokie przemiany w globalnym społeczeństwie i jego dynamice. Nowe narzędzia cyfrowe zmieniają oblicze komunikacji, administracji publicznej, edukacji, konsumpcji, interakcji osobistych i niezliczonych innych aspektów codziennego życia.

Ponadto, technologie wykorzystujące różnorodne algorytmy mogą wydobywać z cyfrowych śladów pozostawionych w *Internecie* dane, które pozwalają kontrolować nawyki mentalne i relacyjne ludzi w celach komercyjnych lub politycznych, często bez ich wiedzy, ograniczając ich świadome korzystanie z wolności wyboru. Rzeczywiście, w przestrzeni takiej jak *sieć*, charakteryzującej się nadmiarem informacji, mogą one kształtować przepływ danych zgodnie z kryteriami wyboru, które nie zawsze są zauważane przez użytkownika.

Musimy pamiętać, że badania naukowe i innowacje technologiczne nie są oderwane od rzeczywistości i „neutralne”[4], lecz podlegają wpływom kulturowym. Jako że, są one w pełni działaniami ludzkimi, to obierane przez nie kierunki odzwierciedlają wybory uwarunkowane wartościami osobistymi, społecznymi i kulturowymi każdej epoki. Odnosi się to także do osiągniętych przez nie rezultatów: właśnie dlatego, że są one wynikiem specyficznego ludzkiego podejścia do otaczającego świata, zawsze mają wymiar etyczny, ściśle związany z decyzjami tych, którzy projektują eksperymenty i ukierunkowują produkcję ku określonym celom.

Dotyczy to również form sztucznej inteligencji. Jak dotąd, w świecie nauki i technologii nie ma jednoznacznej jej definicji. Sam termin, który wszedł już do języka potocznego, obejmuje różnorodność nauk, teorii i technik

mających na celu sprawienie, by maszyny odtwarzały lub naśladowały w swoim działaniu zdolności poznawcze istot ludzkich. Mówienie w liczbie mnogiej o „formach inteligencji” może pomóc podkreślić nade wszystko niemożliwą do pokonania rozbieżność istniejącą między tymi systemami, niezależnie od tego, jak niesamowite i potężne mogą one być, a osobą ludzką: są one w ostatecznym rachunku „fragmentaryczne” w tym sensie, że mogą naśladować lub odtwarzać tylko niektóre funkcje ludzkiej inteligencji. Użycie liczby mnogiej podkreśla również, że te urządzenia, bardzo różniące się między sobą, powinny być zawsze traktowane jako „systemy socjotechniczne”. Istotnie, ich wpływ, niezależnie od technologii podstawowej, zależy nie tylko od ich zaprojektowania, lecz także od celów i interesów tych, którzy je posiadają, i tych, którzy je rozwijają, a także od sytuacji, w jakich są używane.

Sztuczna inteligencja musi być zatem rozumiana jako plejada różnych rzeczywistości i nie możemy zakładać *a priori*, że jej rozwój wniesie dobroczynny wkład dla przyszłości ludzkości i dla pokoju między narodami. Taki pozytywny skutek będzie możliwy tylko wtedy, gdy okażemy się zdolni do odpowiedzialnego działania i poszanowania podstawowych wartości ludzkich, takich jak „inkluzywność, transparentność, bezpieczeństwo, bezstronność, poufność i rzetelność”[5].

Nie wystarczy również zakładać, że osoby projektujące algorytmy i technologie cyfrowe będą działać w sposób etyczny i odpowiedzialny. Konieczne jest wzmocnienie lub, w razie potrzeby, utworzenie organów, powołanych do badania pojawiających się kwestii etycznych i ochrony praw osób korzystających z form sztucznej inteligencji lub znajdujących się pod ich wpływem[6].

Ogromnej ekspansji technologii musi zatem towarzyszyć stosowna formacja w zakresie odpowiedzialności za jej rozwój. Wolność i pokojowe współistnienie są zagrożone wówczas, gdy ludzie ulegają pokusie egoizmu, korzyści osobistej, żądzy zysku i pragnienia władzy. Mamy zatem obowiązek poszerzyć nasze spojrzenie i ukierunkować badania techniczno-naukowe na dążenie do pokoju i dobra wspólnego, w służbie integralnego rozwoju człowieka i wspólnoty[7].

Przyrodzona godność każdej osoby i braterstwo, które łączy nas jako członków jednej rodziny ludzkiej, muszą leżeć u podstaw rozwoju nowych technologii i służyć jako bezdyskusyjne kryteria ich oceny przed ich wykorzystaniem, tak aby postęp cyfrowy mógł odbywać się z poszanowaniem sprawiedliwości i przyczyniać się do sprawy pokoju. Rozwój technologiczny, który nie prowadzi do poprawy jakości życia całej ludzkości, a wręcz przeciwnie, pogłębia nierówności i konflikty, nigdy nie może być uznany za prawdziwy postęp[8].

Sztuczna inteligencja będzie zyskiwać na znaczeniu. Wyzwania jakie stawia mają charakter techniczny, ale także antropologiczny, edukacyjny, społeczny i polityczny. Obiecuje ona na przykład oszczędność nakładów pracy, bardziej wydajną produkcję, sprawniejszy transport i bardziej dynamiczne rynki, a także rewolucję w gromadzeniu, organizacji i procesach weryfikacji danych. Musimy być świadomi zachodzących gwałtownych przemian i zarządzać nimi w sposób, który chroni podstawowe prawa człowieka, z poszanowaniem instytucji i przepisów, które promują jego integralny rozwój. Sztuczna inteligencja powinna służyć najlepszemu ludzkiemu potencjałowi i naszym najbardziej wzniosłym aspiracjom, a nie z nimi konkurować.

### 3. *Technologia przyszłości: maszyny, które same się uczą*

W swoich różnorodnych formach sztuczna inteligencja, oparta na technikach uczenia automatycznego (*machine learning*), choć wciąż znajduje się na etapie pionierskim, już wprowadza znaczące zmiany w tkance społeczeństw, wywierając głęboki wpływ na kultury, zachowania społeczne i budowanie pokoju.

Rozwój, taki jak *machine learning* lub głębokie uczenie (*deep learning*), rodzi pytania, które wykraczają poza sferę technologii i inżynierii, i mają związek ze zrozumieniem, ściśle połączonym z sensem ludzkiego życia, podstawowymi procesami poznawczymi i zdolnością umysłu do dotarcia do prawdy.

Na przykład, zdolność niektórych urządzeń do tworzenia spójnych składniowo i semantycznie tekstów nie jest gwarancją rzetelności. Mówi się, że są one w stanie „oszałamiać”, to znaczy generować stwierdzenia, które na pierwszy rzut oka wydają się wiarygodne, ale w rzeczywistości są bezpodstawne lub zdradzają uprzedzenia.

Stanowi to poważny problem, gdy sztuczna inteligencja jest wykorzystywana w kampaniach dezinformacyjnych, które rozpowszechniają fałszywe wiadomości i prowadzą do rosnącej nieufności wobec środków przekazu. Poufność, posiadanie danych i własność intelektualna to inne obszary, w których omawiane technologie stwarzają poważne zagrożenia, do których dochodzą dalsze negatywne konsekwencje ich niewłaściwego wykorzystania, takie jak: dyskryminacja, ingerencja w procesy wyborcze, szerzenie się modelu społeczeństwa, które monitoruje i kontroluje ludzi, wykluczenie cyfrowe, i pogłębianie się indywidualizmu coraz bardziej oderwanego od zbiorowości. Wszystkie te czynniki mogą podsycać konflikty i utrudniać pokój.

#### 4. Znaczenie ograniczeń w paradygmacie technokratycznym

Nasz świat jest zbyt rozległy, różnorodny i złożony, by można go było w pełni poznać i sklasyfikować. Ludzki umysł nigdy nie będzie w stanie wyczerpać jego bogactwa, nawet z pomocą najbardziej zaawansowanych algorytmów. Nie oferują one bowiem gwarantowanych prognoz przyszłości, a jedynie statystyczne przybliżenia. Nie wszystko można przewidzieć, nie wszystko można obliczyć; w końcu „rzeczywistość przewyższa ideę”[9] i bez względu na to, jak wspaniałe mogą być nasze zdolności obliczeniowe, zawsze pozostanie niedostępna reszta, która wymyka się wszelkim próbom kwantyfikacji.

Ponadto duża ilość danych analizowanych przez sztuczne inteligencje nie jest sama w sobie gwarancją bezstronności. Kiedy algorytmy ekstrapolują informacje, zawsze istnieje ryzyko, że je wypaczają, powielając niesprawiedliwości i uprzedzenia środowisk, z których pochodzą. Im bardziej stają się szybsze i złożone, tym trudniej zrozumieć, dlaczego wyprodukowały określony rezultat.

Maszyny inteligentne mogą wykonywać przypisane im zadania z coraz większą wydajnością, ale cel i znaczenie ich działań będą nadal określane lub umożliwiane przez ludzi posiadających własny świat wartości. Istnieje ryzyko, że kryteria stojące za niektórymi wyborami staną się mniej jasne, że odpowiedzialność za podejmowanie decyzji zostanie ukryta, a producenci mogą uchylać się od obowiązku działania dla dobra wspólnoty. W pewnym sensie sprzyja temu system technokratyczny, który sprzymierza ekonomię z technologią i nadaje przywilej dla kryterium wydajności, ignorując wszystko, co nie jest związane z jego doraźnymi korzyściami[10].

Powinno to nas skłonić do refleksji nad aspektem, który jest bardzo często pomijany w dzisiejszej mentalności technokratycznej i wydajnościowej, jako decydujący dla rozwoju osobistego i społecznego: „poczucie ograniczeń”. Istocie ludzkiej, z definicji śmiertelnej, gdy myśli o przekroczeniu, dzięki technologii, wszelkich ograniczeń, grozi, iż będąc ogarniętą obsesją kontrolowania wszystkiego, utraci kontrolę nad sobą; w poszukiwaniu absolutnej wolności, wpadnie w spiralę dyktatury technologicznej. Rozpoznanie i zaakceptowanie własnej ograniczoności jako stworzenia, jest dla człowieka niezbędnym warunkiem osiągnięcia pełni, a raczej przyjęcia jej w darze. Natomiast w ideologicznym kontekście paradygmatu technokratycznego, ożywionego prometejskim założeniem samowystarczalności, nierówności mogą rosnać nieproporcjonalnie, a wiedza i bogactwo gromadzić się w rękach nielicznych, z poważnym zagrożeniem dla społeczeństw demokratycznych i pokojowego współistnienia[11].

#### 5. Gorące tematy dla etyki

W przyszłości, wiarygodność osoby ubiegającej się o kredyt, przydatność danej osoby do pracy, prawdopodobieństwo recydywy osoby skazanej, lub prawo do otrzymania azylu politycznego, lub pomocy społecznej, mogą być determinowane przez systemy sztucznej inteligencji. Brak zróżnicowanych poziomów mediacji, które wprowadzają te systemy, jest szczególnie narażony na formy uprzedzeń i dyskryminacji: błędy systemowe mogą się łatwo mnożyć, powodując nie tylko niesprawiedliwość w indywidualnych przypadkach, ale także, poprzez efekt domina, rzeczywiste formy nierówności społecznej.

Co więcej, czasami formy sztucznej inteligencji wydają się zdolne do wpływania na decyzje jednostek, poprzez z góry określone opcje związane z bodźcami i środkami odstraszającymi lub poprzez systemy regulujące osobiste wybory w oparciu o organizację informacji. Te formy manipulacji lub kontroli społecznej wymagają uwagi i starannego nadzoru, oraz pociągają za sobą wyraźną odpowiedzialność prawną ze strony producentów, tych, którzy je stosują, oraz władz rządowych.

Poleganie na procesach automatycznych, które kategoryzują jednostki, na przykład poprzez wszechobecne wykorzystanie nadzoru lub przyjęcie systemów kredytu społecznego, może mieć również głębokie reperkusje dla tkanki obywatelskiej, ustanawiając niewłaściwe klasyfikacje wśród obywateli. Te sztuczne procesy klasyfikacji mogą również prowadzić do konfliktów władzy, ponieważ dotyczą nie tylko odbiorców wirtualnych, ale także osób z krwi i kości. Fundamentalny szacunek dla godności ludzkiej postuluje odrzucenie tego, aby utożsamiano wyjątkowość osoby ze zbiorem danych. Nie można pozwolić algorytmom: określać sposobu, w jaki rozumiemy prawa człowieka, odkładać na bok podstawowych wartości współczucia, miłosierdzia i przebaczenia lub eliminować możliwość, aby jednostka się zmieniała i pozostawiła przeszłość za sobą.

W tym kontekście nie możemy nie rozważyć wpływu nowych technologii w sferze zatrudnienia: prace, które kiedyś były wyłączną domeną siły roboczej, są szybko wchłaniane przez przemysłowe zastosowania sztucznej inteligencji. Również w tym przypadku istnieje znaczne ryzyko nieproporcjonalnej przewagi nielicznych, kosztem zubożenia wielu. Poszanowanie godności pracowników i znaczenie zatrudnienia dla dobrobytu ekonomicznego osób, rodzin i społeczeństw, bezpieczeństwo pracy i sprawiedliwe płace, powinny być wysokim priorytetem dla wspólnoty międzynarodowej, ponieważ te formy technologii przenikają coraz głębiej do miejsc pracy.

## 6. Czy przekazujemy miecze na lemieszce?

W dzisiejszych czasach, patrząc na otaczający nas świat, nie sposób uciec od poważnych kwestii etycznych związanych z przemysłem zbrojeniowym. Możliwość prowadzenia operacji wojskowych za pomocą systemów zdalnie sterowanych doprowadziła do zmniejszenia percepcji zniszczeń, jakie one powodują i odpowiedzialności za ich użycie, przyczyniając się do jeszcze zimniejszego i bardziej oderwanego podejścia do ogromnej tragedii wojny. Badania nad nowymi technologiami w dziedzinie tak zwanych „śmiercionośnych autonomicznych systemów uzbrojenia”, w tym nad wykorzystaniem wojennym sztucznej inteligencji, stanowią poważny problem etyczny. Autonomiczne systemy uzbrojenia nigdy nie mogą być podmiotami odpowiedzialnymi moralnie: unikalna ludzka zdolność do moralnej oceny i etycznego podejmowania decyzji jest czymś więcej niż złożonym zestawem algorytmów, a zdolności tej nie można sprowadzać do programowania maszyny, która, choć „inteligentna”, nadal jest maszyną. Z tego powodu konieczne jest zapewnienie odpowiedniego, znaczącego i spójnego ludzkiego nadzoru nad systemami uzbrojenia.

Nie możemy również lekceważyć potencjalnego ryzyka, że wyrafinowana broń wpadnie w niepowołane ręce, ułatwiając na przykład ataki terrorystyczne lub ingerencje mające na celu destabilizację prawowitych instytucji rządowych. Krótko mówiąc, świat naprawdę nie potrzebuje nowych technologii przyczyniających się do nieuczciwego rozwoju rynku i handlu bronią, promujących szaleństwo wojny. Czyniąc w ten sposób, nie tylko inteligencja, ale samo serce człowieka będzie narażone na ryzyko, że stanie się coraz bardziej „sztuczne”. Najbardziej zaawansowane aplikacje techniczne nie powinny być wykorzystywane do ułatwiania brutalnego rozwiązywania konfliktów, lecz do przygotowywania drogi do pokoju.

Z bardziej pozytywnej perspektywy, gdyby sztuczna inteligencja została wykorzystana do promowania integralnego rozwoju człowieka, mogłaby wprowadzić znaczące innowacje w rolnictwie, edukacji i kulturze, poprawić standardy życia całych narodów i ludów oraz przyczynić się do wzrostu ludzkiego braterstwa i przyjaźni społecznej. Ostatecznie, to, w jaki sposób wykorzystamy ją do włączenia najmniejszych, czyli naszych najsłabszych i najbardziej potrzebujących braci i siostr, jest miarą ukazującą nasze człowieczeństwo.

Ludzkie spojrzenie i pragnienie lepszej przyszłości dla naszego świata prowadzą do potrzeby interdyscyplinarnego dialogu, mającego na celu etyczny rozwój algorytmów – *algor-etykę* – w którym wartości kierowałyby drogami nowych technologii[12]. Kwestie etyczne powinny być brane pod uwagę od samego początku badań, a także na etapie testowania, projektowania, produkcji, dystrybucji i wprowadzenia na rynek. Jest to podejście etyki projektowania, w którym instytucje edukacyjne i decydenci mają do odegrania istotną rolę.

## 7. Wyzwania dla edukacji

Rozwój technologii, która szanuje i służy ludzkiej godności, ma wyraźne implikacje dla instytucji edukacyjnych i

świata kultury. Zwielokrotniając możliwości komunikacji, technologie cyfrowe umożliwiły spotkania na nowe sposoby. Istnieje jednak potrzeba stałej refleksji nad rodzajem relacji, do których nas kierują. Młodzi dorastają w środowiskach kulturowych przenikniętych technologią, co nie może nie budzić wątpliwości w zakresie metod nauczania i formacji.

Edukacja w zakresie korzystania z form sztucznej inteligencji powinna mieć na celu przede wszystkim promowanie krytycznego myślenia. Konieczne jest, aby użytkownicy w każdym wieku, a zwłaszcza ludzie młodzi, rozwijali umiejętność świadomego i wnikliwego korzystania z danych i treści gromadzonych w *sieci* lub tworzonych przez systemy sztucznej inteligencji. Szkoły, uniwersytety i towarzystwa naukowe są wezwane do pomocy studentom i profesjonalistom w uwzględnieniu społecznych i etycznych aspektów rozwoju i wykorzystania technologii.

Formacja, w zakresie korzystania z nowych narzędzi komunikacji, powinna uwzględniać nie tylko dezinformację, fałszywe wiadomości (tzw. *fake news*), ale także niepokojące odradzanie się „dawnych obaw, które [...] potrafiły się ukryć i umocnić swoją pozycję za nowymi technologiami”[13]. Niestety, po raz kolejny musimy walczyć z „pokusą tworzenia kultury murów, wznoszenia murów [...], aby uniemożliwić [...] spotkanie z innymi kulturami, z innymi ludźmi” [14], oraz rozwój pokojowego i braterskiego współistnienia.

#### 8. Wyzwania dla rozwoju prawa międzynarodowego

Globalny zasięg sztucznej inteligencji jasno pokazuje, że oprócz odpowiedzialności suwerennych państw za regulowanie jej wykorzystania we własnym kraju, decydującą rolę w zawieraniu porozumień wielostronnych oraz koordynowaniu ich stosowania i wdrażania mogą odegrać organizacje międzynarodowe[15]. W związku z tym, zachęcam Wspólnotę Narodów, aby – zjednoczona – pracowała w celu przyjęcia wiążącego traktatu międzynarodowego, regulującego rozwój i wykorzystanie sztucznej inteligencji w jej różnorodnych formach. Celem regulacji powinno być, oczywiście, nie tylko zapobieganie złym praktykom, ale także zachęcanie do dobrych praktyk, stymulowanie nowych i kreatywnych koncepcji oraz ułatwianie inicjatyw osobistych i zbiorowych[16].

W ostatecznym rachunku, w poszukiwaniu modeli regulacyjnych, które mogłyby zapewnić twórcom technologii cyfrowych pewien przewodnik etyczny, konieczne jest wskazanie ludzkich wartości, które powinny leżeć u podstaw wysiłków społeczeństw na rzecz sformułowania, przyjęcia i zastosowania niezbędnych ram prawnych. Prace redakcyjne nad wytycznymi w zakresie etyki, dotyczącymi produkcji form sztucznej inteligencji, nie mogą pomijać głębszych kwestii odnoszących się do sensu ludzkiego istnienia, ochrony podstawowych praw człowieka oraz dążenia do sprawiedliwości i pokoju. Ten proces etycznego i prawnego rozeznania może okazać się cenną okazją do wspólnej refleksji nad rolą, jaką technologia powinna odgrywać w naszym życiu indywidualnym i wspólnotowym, oraz w jaki sposób jej wykorzystanie może przyczynić się do stworzenia bardziej sprawiedliwego i humanitarnego świata. Z tego powodu, w debatach na temat regulacji sztucznej inteligencji powinny być brane pod uwagę głosy wszystkich zainteresowanych stron, w tym ubogich, wykluczonych i innych, którzy często pozostają nieusłyszani w globalnych procesach decyzyjnych.

\* \* \* \* \*

Mam nadzieję, że ta refleksja zachęci, do uczynienia wszystkiego, aby postępy w rozwoju form sztucznej inteligencji w ostateczności służyły sprawie ludzkiego braterstwa i pokoju. Nie jest to odpowiedzialność nielicznych, lecz całej rodziny ludzkiej. Pokój jest bowiem owocem relacji, które uznają i akceptują drugiego człowieka w jego niezbywalnej godności, a także owocem współpracy i zaangażowania w dążeniu do integralnego rozwoju wszystkich osób i wszystkich narodów.

Na początku Nowego Roku modłę się, aby szybki rozwój form sztucznej inteligencji nie powiększył zbyt wielu nierówności i niesprawiedliwości już obecnych na świecie, ale przyczynił się do zakończenia wojen i konfliktów oraz złagodzenia wielu form cierpienia, które dotyczą rodzinę ludzką. Oby chrześcijanie, wyznawcy różnych religii, oraz mężczyźni i kobiety dobrej woli, zgodnie współpracowali, żeby wykorzystać szanse i sprostać wyzwaniom stawianym przez rewolucję cyfrową, oraz przekazać przyszłym pokoleniom bardziej solidarny,

sprawiedliwy i pokojowy świat.

Watykan, 8 grudnia 2023 r.

FRANCISZEK

---

[1] Sobór Wat. II, Konstytucja duszp. o Kościele w świecie współczesnym, *Gaudium et spes*, 33.

[2] *Tamże*, 57.

[3] Por. Enc. *Laudato si'* (24 maja 2015), 104.

[4] *Tamże*, 114.

[5] *Udienza ai partecipanti all'Incontro "Minerva Dialogues"* (27 marzo 2023).

[6] Por. *tamże*.

[7] Por. *Messaggio al Presidente Esecutivo del "World Economic Forum" a Davos-Klosters* (12 gennaio 2018).

[8] Por. Enc. *Laudato si'*, 194; *Discorso ai partecipanti al Seminario "Il bene comune nell'era digitale"* (27 settembre 2019).

[9] Adhort. apost. *Evangelii gaudium* (24 listopada 2013), 233.

[10] Por. Enc. *Laudato si'*, 54.

[11] Por. *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita* (28 febbraio 2020).

[12] Por. *tamże*.

[13] Enc. *Fratelli tutti* (3 października 2020), 27.

[14] *Tamże*.

[15] Por. *tamże*, 170-175.

[16] Por. Enc. *Laudato si'*, 177.

[01927-PL.01] [Testo originale: Italiano]

Traduzione in lingua araba

أبابل ةس ادق ةلاس ر

س س نرف

2024 ريان ي/ينأأال نوناك نم لوالا

مأسل او يعان طصالا ءاكذلا

في بداية السنة الجديدة، زمن النعمة الذي يمنحه الله لكل واحد منا، أود أن أتوجه إلى شعب الله، وإلى شعوب العالم، ورؤساء الدول والحكومات، وممثلي الديانات المختلفة والمجتمع المدني، وإلى جميع الرجال والنساء في زمننا، لأقدم للجميع أمنية سلام شامل.

## 1. تقدم العلم والتكنولوجيا كطريق للسلام

يوكد الكتاب المقدس أن الله أعطى الإنسان روحه ليملاه "مهارةً وفهماً ومعرفةً بجمیع الصناعات" (خروج 35، 31). الفهم أو الذكاء هو تعبير عن الكرامة التي أعطانا إياها الخالق، الذي خلقنا على صورته ومثاله (راجع تكوين 1، 26)، ومكنا من أن نجيب على محبته بالحرية والمعرفة. والعلم والتكنولوجيا يظهران بصورة خاصة هذه الطيبة العلائقية الأساسية للذكاء البشري: فهما تتاجان استثنائياً لقدرته الإبداعية.

في الدستور الرعائي فرح ورجاء، أكد المجمع الفاتيكاني الثاني هذه الحقيقة، وصرح أن "الإنسان سعى دائماً لتطوير حياته بعمله وذكائه"[1]. عندما يسعى البشر، "بالأدوات التقنية"، لكي يجعلوا الأرض "مسكناً لائقاً بالعائلة البشرية كلها"[2]، فإنهم يعملون وفقاً لتدبير الله ويتعاونون مع إرادته لإكمال الخليفة ونشر السلام بين الشعوب. وتقدم العلم والتكنولوجيا أيضاً، يؤدي إلى تحسين الإنسان وتغيير العالم، بقدر مساهمتهما في تحسين نظام المجتمع البشري، وفي زيادة الحرية والشركة الأخوية.

ونحن بحق نفرح ونشعر بالشكر وعرقان الجميل للإنجازات غير العادية التي حققها العلم والتكنولوجيا، والتي بفضلها وجد علاج لعدد لا يحصى من الشرور التي ابتليت بها الحياة البشرية وتسببت في آلام كبيرة. وفي الوقت نفسه، فإن التقدم التقني والعلمي، الذي يجعل من الممكن ممارسة سيطرة غير مسبوق حتى الآن على الواقع، يضع مجموعة واسعة من الإمكانيات في يدي الإنسان، والتي قد يكون بعضها خطراً على بقاء الإنسان وعلى بيتنا المشترك[3].

ولذلك فإن التقدم الملحوظ الذي حققته تكنولوجيات المعلومات الجديدة، وخاصة في المجال الرقمي، إنما هي فرص مدهشة، وفي الوقت نفسه مجازفة خطيرة، ولها آثار جدية في السعي لتحقيق العدل والوئام بين الشعوب. ولذلك لا بد من طرح بعض الأسئلة الملحة. ما هي العواقب في المدى القريب والبعيد للتكنولوجيات الرقمية الجديدة؟ وماذا سيكون تأثيرها على حياة الأفراد والمجتمع وعلى الاستقرار الدولي والسلام؟

## 2. مستقبل الذكاء الاصطناعي بين الوعود والمخاطر

التقدم في تكنولوجيا المعلومات وتطور التكنولوجيات الرقمية بدأ يحدث، في العقود الأخيرة، تغييرات عميقة في المجتمع العالمي وديناميكياته. والأدوات الرقمية الجديدة أخذت تغير وجه الاتصالات والإدارة العامة والتعليم والاستهلاك والعلاقات بين الأشخاص، وجوانب أخرى لا حصر لها في الحياة اليومية.

وكذلك، التغيرات التي تستخدم عدداً كبيراً من الخوارزميات يمكنها أن تستخرج، من الآثار الرقمية المتبقية على "الإنترنت"، البيانات التي تسمح بالتحكم في عادات الأشخاص العقلية والعلائقية لأغراض تجارية أو سياسية، غالباً دون علمهم، ما يحد من ممارستهم الواعية لحرية الاختيار. في الواقع، في مساحة مثل الشبكة المعلوماتية العالمية (web)،

علينا أن نتذكر أن البحث العلمي والابتكارات التكنولوجية ليست منزوعة من الواقع ولا هي "حيادية" [4]، ولكنها تخضع للمؤثرات الثقافية. ولكونها أنشطة إنسانية بكل معنى الكلمة، فإن الاتجاهات التي تتخذها تعكس اختيارات متأثرة بالقيم الشخصية والاجتماعية والثقافية لكل عصر. والأمر نفسه ينطبق على النتائج التي تحققها: فلأنها نتيجة لمقاربات إنسانية في البيئة المحيطة بها، لها دائماً بُعد أخلاقي، يرتبط ارتباطاً وثيقاً بقرارات الذين يخططون للتجربة ويوجهون الإنتاج نحو أهداف خاصة.

وهذا ينطبق أيضاً على أشكال الذكاء الاصطناعي. حتى الآن، لا يوجد تعريف موحد له في عالم العلم والتكنولوجيا. المصطلح نفسه، الذي دخل الآن في اللغة العامة، يشمل مجموعة متنوعة من العلوم والنظريات والتقنيات التي تهدف إلى جعل الآلات تنتج أو تقلد القدرات المعرفية للبشر في أدائها. الحديث بصيغة الجمع عن "أشكال الذكاء" يمكن أن يساعد في المقام الأول على التأكيد على الفجوة التي لا يمكن ردمها والتي توجد بين هذه الأنظمة، مهما كانت عجيبة وقوية، وبين الإنسان: فهي في نهاية المطاف "مجزأة"، بمعنى أنها تستطيع فقط تقليد أو إعادة إنتاج بعض وظائف الذكاء البشري. استخدام صيغة الجمع يبين أن هذه الأدوات، المختلفة كثيراً بعضها عن بعض، يجب اعتبارها دائماً "أنظمة اجتماعية تقنية". وفي الواقع، فإن تأثيرها، بغض النظر عن التكنولوجيا الأساسية، لا يعتمد على التصميم فحسب، بل يعتمد أيضاً على أهداف ومصالح الذين يمتلكونها والذين يقومون بتطويرها، وكذلك على الحالات التي تُستخدم فيها.

ولذلك، يجب أن يفهم الذكاء الاصطناعي على أنه كوكبة من الحقائق المختلفة، ولا يمكننا أن نفترض بدهاءة أن تطوره سيقدم مساهمة مفيدة لمستقبل البشرية وللسلام بين الشعوب. ولن تكون هذه النتيجة الإيجابية ممكنة إلا إذا أثبتنا أننا قادرون على التصرف بمسؤولية واحترام القيم الإنسانية الأساسية مثل "الشمولية والشفافية والأمن والعدالة والسرية والثقة" [5].

ولا يكفي حتى أن نفترض التزام الذين يصممون الخوارزميات والتقنيات الرقمية بالتصرف بطريقة أخلاقية ومسؤولة. ينبغي تعزيز، أو إذا لزم الأمر، إنشاء هيئات مسؤولة عن دراسة القضايا الأخلاقية المترتبة عليها وحماية حقوق الذين يستخدمون بعض أشكال الذكاء الاصطناعي أو يتأثرون بها [6].

ولذلك فإن التوسع الهائل في التكنولوجيا يجب أن يكون مصحوباً بالتدريب الكافي على المسؤولية في تطويرها. الحرية والعيش معاً بسلام يتعرضان للتهديد عندما يستسلم البشر لتجارب الأناية والمصلحة الشخصية والجشع في الربح والتعطش إلى السلطة. ولذلك، يقع على عاتقنا واجب توسيع الرؤية وتوجيه البحث العلمي والتقني إلى تحقيق السلام والخير العام، في خدمة التنمية المتكاملة للإنسان والمجتمع [7].

الكرامة الجوهرية لكل شخص والأخوة التي تربطنا أعضاءً في الأسرة البشرية الواحدة يجب أن تكون أساساً لتطوير التكنولوجيات الجديدة، فتكون بمثابة معايير لا جدال فيها لتقييمها قبل استخدامها، حتى يتمكن التقدم الرقمي من أن يتحقق مع احترام العدل ومع المساهمة في قضية السلام. التطورات التكنولوجية التي لا تؤدي إلى تحسين نوعية حياة البشرية جمعاء، بل عكس ذلك تؤدي إلى تفاقم عدم المساواة والصراعات، لا يمكن اعتبارها تقدماً حقيقياً [8].

ستزداد أهمية الذكاء الاصطناعي. والتحديات التي تثيرها هي تحديات تقنية، ولكنها أيضاً أثروبولوجية وتربوية واجتماعية وسياسية. فهي تعد، مثلاً، بتوفير الجهد، وإنتاج أكثر فعالية، ووسائل نقل أكثر راحة، وحركة متزايدة في الأسواق، فضلاً عن ثورة في عمليات جمع البيانات وتنظيمها والتحقق منها. وعلينا أن ندرك التحولات السريعة المستمرة وأن نديرها بطريقة تصون حقوق الإنسان الأساسية، وتحترم المؤسسات والقوانين التي تعزز التنمية البشرية المتكاملة. الذكاء الاصطناعي يجب أن يخدم أفضل الإمكانيات البشرية وأسمى تطلعاتنا، ولا يتنافس معها.

### 3. تكنولوجيا المستقبل: آلات تتعلم بنفسها

الذكاء الاصطناعي، بأشكاله المتعددة، القائم على تقنيات التعلم التلقائي (التعلم الآلي)، رغم أنه لا يزال في مرحلة رائدة، فإنه بدأ يدخل تغييرات ملحوظة في نسيج المجتمعات، وحدث تأثيراً عميقاً في الثقافات والسلوكيات الاجتماعية وبناء السلام.

التطورات مثل التعلم الآلي أو التعلم العميق تطرح أسئلة تفوق مجالات التكنولوجيا والهندسة، ولها صلة بفهم مرتبط ارتباطاً وثيقاً بمعنى الحياة البشرية وعمليات المعرفة الأساسية وقدرة العقل على الوصول إلى الحقيقة.

قدرة بعض الأجهزة على إنتاج نصوص متماسكة ومتناسقة من حيث المعنى، على سبيل المثال، ليس ضماناً لثيق بها. يقال إنهم يستطيعون أن "يحدثوا وهمًا"، أي أن يصدروا عبارات تبدو للوهلة الأولى معقولة، ولكن في الواقع لا أساس لها من الصحة أو يمكن أن تدل على أحكام مسبقة. وهذه مشكلة خطيرة عندما يُستخدم الذكاء الاصطناعي في حملات التضليل التي تشر أخباراً مزيفة وتؤدي إلى تزايد عدم الثقة بوسائل الإعلام. السرية وحيازة البيانات والملكية الفكرية هي مجالات أخرى تشكل فيها هذه التكنولوجيات مخاطر جسيمة، يضاف إليها المزيد من العواقب السلبية المرتبطة باستخدامها بصورة غير سليمة، مثل التفرقة، والتدخل في العمليات الانتخابية، والانحياز إلى مجتمع يراقب ويتحكم بالناس، وحرمان الاستخدام الرقمي، وتفاقم نزعة فردية تسبب انفصلاً متزايداً عن الجماعة. كل هذه العوامل تهدد بتغذية الصراعات وعرقلة السلام.

### 4. حسّ المحدودية في النموذج التكنولوجي

عالمنا واسع جداً، ومتنوع ومعقد فلا يمكن أن نعرفه معرفة كاملة وأن نصنّفه. لا يمكن للعقل البشري أن يدرك بصورة كاملة غنى العالم، ولا حتى بمساعدة أكثر الخوارزميات تقدماً. في الواقع، هذه الخوارزميات لا تقدم لنا توقعات مضمونة عن المستقبل، بل إحصائيات تقريبية فقط. لا يمكننا التنبؤ بكل شيء، ولا يمكننا أن نحسب كل شيء، لأنه في النهاية "الواقع أسمى من الفكرة" [9]، ومهما بلغت قدرتنا الحاسوبية، سيكون هناك دائماً أمور متبقية لا يمكننا أن نصّل إليها، فهي تعصى على كل محاولة لإخضاعها لقياساتنا.

علاوة على ذلك، كمية البيانات الكبيرة التي يحلّلها الذكاء الاصطناعي ليست بحد ذاتها ضماناً للحيادية. عندما تخلق الخوارزميات بين المعلومات عند استخدامها، فإنها تهدد بتشويهها، وتكرّر المظالم والأحكام المسبقة في البيانات التي تنشأ فيها. وكلما زادت السرعة، وزاد التعقيد فيها، كلما زادت صعوبة الفهم لماذا نتجت هذه النتيجة المحددة.

يمكن للآلات الذكية أن تؤدي المهام الموكولة إليها بكفاءة متزايدة، لكن، الهدف والمعنى لعملياتها، سيستمرّ الإنسان هو الذي يحددها أو يفعلها، والإنسان له عالم خاص به من القيم. يكمن الخطر في أن المعايير لبعض الاختيارات تصبح أقل وضوحاً، وتغيب المسؤولية في اتخاذ القرار، ويتخلّى المنتجون عن الالتزام بالعمل من أجل خير الجماعة. بمعنى ما، يعزز هذه الإمكانية النظام التكنولوجي، الذي يربط الاقتصاد والتكنولوجيا، ويفضل معيار الفعالية في الإنتاج، فيميل إلى تجاهل كل ما لا يرتبط بمصالحه المباشرة [10].

هذا الأمر يجب أن يجعلنا نفكر في جانبٍ كثيراً ما تتجاهله في العقلية التكنولوجية والموجهة نحو الإنتاج، التي نعيشها في يومنا هذا. وهو أمر حاسم للتنمية الشخصية والاجتماعية، وهو: "حسّ المحدودية". في الواقع، الإنسان، الذي هو كائنٌ فإن بتعريفه، ويفكر في تجاوز كل الحدود بفضل التكنولوجيا، يوشك أن يفقد السيطرة على نفسه، في هوسه لإخضاع كل شيء لسيطرته. وفي بحثه عن الحرية المطلقة، يُخاطر أن يقع في دوامة "الديكتاتورية التكنولوجية". الاعتراف بحدودنا كخليفة والقبول بها، هو شرط لا غنى عنه للإنسان لكي يحقق، أو بالأحرى، لكي يقبل الكمال

## 5. مواضع صارخة في مجال الأخلاق

في المستقبل، مصداقية الشخص الذي يطلب قرصاً مالياً، وملاءمة شخص ما لوظيفة، وإمكانية وقوع الشخص المحكوم عليه في الجرم ثانية، أو الحق في الحصول على اللجوء السياسي أو المساعدة الاجتماعية، كل ذلك يمكن أن تصبح أنظمة الذكاء الاصطناعي هي التي تحدده. هذه الأنظمة تلغي مختلف مستويات الوساطة، وتعرض للوقوع في أشكال من التحيز والتفرقة: إذ يمكن أن تتضاعف بسهولة الأخطاء المرتبطة بنظام، فتنتج ظلاماً ليس فقط في حالات فردية، بل أيضاً أخطاء متلاحقة مثل تابع وقوع حجارة الدومينو، وتسبب أشكالاً حقيقية من عدم المساواة الاجتماعية.

بالإضافة إلى ذلك، تبدو أحياناً أشكال الذكاء الاصطناعي قادرة على أن تؤثر على قرارات الأفراد من خلال خيارات محددة مسبقاً ومرتبطة بمحفّزات وإقناع بالعدول عن الموقف، أو من خلال أنظمة تتحكم بالخيارات الشخصية بناء على كيفية تنظيم المعلومات. هذه الأشكال من التلاعب والخداع أو السيطرة الاجتماعية، تتطلب تبهها وتدقيقاً حذراً، وتقع فيها مسؤولية قانونية واضحة على المنتجين، والذين يستخدمونها والسلطات الحكومية.

الاعتماد على عمليات أوتوماتيكية تصنف الأفراد، مثلاً من خلال استخدام مفرط للرقابة أو اعتماد أنظمة الائتمان الاجتماعي، يمكن أن يكون له تداعيات عميقة، وأيضاً على النسيج المدني، ويحدد تصنيفات غير صحيحة بين المواطنين. ويمكن أن تؤدي عمليات التصنيف المصطنعة هذه إلى صراعات على السلطة أيضاً، ليس فقط بين الأشخاص الافتراضيين، بل بين الأشخاص الحقيقيين أيضاً. الاحترام الأساسي للكرامة الإنسانية يتطلب منا رفض أن يتم تحديد فرادة الشخص من خلال مجموعة من البيانات. يجب ألا نسمح للخوارزميات أن تحدد الطريقة التي فيها نفهم حقوق الإنسان، وأن تضع جانباً القيم الأساسية للرفاه والرحمة والغفران، أو إلغاء إمكانية أن يتغير الفرد وأن يترك ماضيه وراءه.

في هذا السياق، لا يسعنا إلا أن نأخذ بعين الاعتبار تأثير التكنولوجيات الجديدة في مجال العمل: الأعمال التي كانت في وقت من الأوقات حكرًا على العمل البشري، سيطرت عليها بسرعة التطبيقات الصناعية للذكاء الاصطناعي. في هذه الحالة أيضاً، يوجد خطر كبير لتحقيق فائدة غير متكافئة لعدد قليل من الناس، على حساب كثيرين يعرضون للفقر. احترام كرامة العمال وأهمية العمل لتحقيق الرفاه الاقتصادي للأشخاص والعائلات والمجتمعات، والأمن الوظيفي والأجور العادلة، كل ذلك عليه أن يشكل أولوية عليا للمجتمع الدولي، فيما تتغلغل أشكال التكنولوجيا هذه بشكل أعمق دائماً في أماكن العمل.

## 6. هل سنحول السيوف إلى سكاك للحراثة؟

في هذه الأيام، ونحن ننظر إلى العالم من حولنا، لا يمكننا أن نتهرب من القضايا الأخلاقية الخطيرة المرتبطة بقطاع التسلح. إمكانية إجراء عمليات عسكرية من خلال أنظمة التحكم عن بعد أدت إلى عدم رؤية الدمار الذي تسببه هذه الأنظمة، وإلى عدم الإحساس بمسؤولية استخدامها، فتتظر بمزيد من البرود واللامبالاة إلى مأساة الحرب الهائلة. الأبحاث في التقنيات التي تنشأ في القطاع المسمى "بأنظمة الأسلحة الفتاكة العاملة بصورة آلية"، بما في ذلك استخدام الذكاء الاصطناعي العسكري، هو سبب خطير يثير القلق على الصعيد الأخلاقي. لا يمكن لأنظمة الأسلحة الآلية أن تكون هي مسؤولة أخلاقياً: الإنسان وحده له القدرة على الحكم الأخلاقي واتخاذ القرار الأخلاقي، والإنسان أكثر بكثير من مجرد مجموعة معقدة من الخوارزميات. ومن ثم، لا يمكن حصر هذه القدرة الأخلاقية في برمجة آلية، فهي مهما كانت "ذكية"، تبقى دائماً آلة. لهذا السبب، لا بد من أن نضمن رقابة بشرية مناسبة، ومنسجمة ولها أهميتها، على أنظمة الأسلحة.

ولا نغدر أن تتجاهل إمكانيّة وقوع الأسلحة المتطوّرة في الأيدي الخطأ، فتسهّل على سبيل المثال، الهجمات الإرهابية أو التّدخلات التي تهدف إلى زعزعة استقرار أنظمة حكوميّة شرعيّة. باختصار، العالم لا يحتاج إلى تقنيّات جديدة لتطوير آثم للسّوق، ولتجارة الأسلحة، تعزّز جنون الحرب. إن فعلنا ذلك، ليس الذّكاء وحده، بل قلب الإنسان نفسه، يصبح عرضة لأن يصير "اصطناعياً". يجب ألاّ نستخدم أكثر التطبيقات التّقنيّة تطوراً لتسهيل حلّ الصّراعات بالعنف، بل لتمهيد الطّرق إلى السّلام.

من منظور أكثر إيجابيّة، لو تمّ استخدام الذّكاء الاصطناعيّ لتعزيز التّمية البشريّة المتكاملة، لأمكنه أن يقدّم ابتكارات مهمّة في الزراعة والتّعليم والثّقافة، وتحسين في مستوى المعيشة لأمم وشعوب بأكملها، وازدادت الأخوة الإنسانيّة والصّدقة الاجتماعيّة. في النّهاية، الطّريقة التي بها نستخدمه لنشمل الآخرين، أيّ الإخوة والأخوات الأكثر ضعفاً واحتياجاً، هي المقياس التي تُظهر إنسانيّتنا.

النّظرة الإنسانيّة والرّغبة في مستقبل أفضل لعالمنا، يدفعان إلى ضرورة الحوار بين مختلف العلوم والتخصّصات، بهدف إلى التطوير الأخلاقيّ للخوارزميّات - أخلاقيّات الخوارزميّات - فيه توجّه القيم مسارات التّقنيّات الجديدة [12]. يجب أن تؤخذ القضايا الأخلاقيّة بعين الاعتبار منذ بداية البحث، وفي مراحل الاختبار أيضاً، والتصميم والتصنيع والتوزيع والتسويق. هذا هو النّهج الأخلاقيّ للتصميم، الذي فيه يكون للمؤسّسات التّربويّة وصّناع القرار دور أساسيّ يمارسونه.

## 7. تحديات التّربية

تطوير التّكنولوجيا التي تحترم وتخدم الكرامة الإنسانيّة، له ارتباطات واضحة بالمؤسّسات التّربويّة وعالم الثّقافة. بمضاعفة إمكانيّات الاتّصال، سمحت التّقنيّات الرّقميّة بأن نلتقي بطرق جديدة. مع ذلك، لا زلنا بحاجة لأن نتأمّل باستمرار في نوع العلاقات التي توجّهنا إليها هذه التّقنيّات. الشّباب اليوم ينمون في بيئات ثقافيّة مشبعة بالتّكنولوجيا، وهذا الأمر لا يمكن إلاّ أن يثير تساؤلات حول أساليب التّدرّس والتّنشئة.

التّربية على استخدام أشكال الذّكاء الاصطناعيّ يجب أن تهدف، قبل كلّ شيء، إلى تعزيز التّفكير النّقدي. من المهمّ للمستخدمين من جميع الأعمار، وخاصة الشّباب، أن يطوروا قدرتهم على التّمييز في استخدام البيانات والمحتويات المجموعة على الشّبكات المعلوماتيّة العالميّة (web) أو التي أنتجتها أنظمة الذّكاء الاصطناعيّ. المدارس والجامعات والجمعيات العلميّة مدعوّة إلى أن تساعد الطّلاب والمختصّين ليتبنوا الجوانب الاجتماعيّة والأخلاقيّة لتطوير واستخدام التّكنولوجيا.

التّنشئة على استخدام أدوات الاتّصال الجديدة يجب أن تأخذ بعين الاعتبار، ليس فقط تشويه المعلومات والأخبار المزيفة، بل أيضاً العودة المقلقة إلى "مخاوف قديمة [...] استطاعت أن تختبئ وتتطور خلف التّقنيّات الجديدة" [13]. للأسف، نجد أنفسنا مرّة أخرى مضطّرين لأن نواجه "الميل لإقامة ثقافة الجدران، ورفع الجدران، لمنع اللقاء مع الثّقافات الأخرى، ومع الآخرين" [14] وتطوير العيش معاً بسلام وأخوة.

## 8. تحديات تطوير القانون الدّوليّ

انتشار الذّكاء الاصطناعيّ العالميّ يوضّح، إلى جانب مسؤوليّة الدّول ذات السّيادة لتنظيم استخدامه داخليّاً، دور المنظّمات الدّولية التي يمكنها أن تلعب دوراً حاسماً في التّوصّل إلى اتّفاقيّات متعدّدة الأطراف وتنسيق تطبيقها وتنفيذها [15]. وفي هذا الصّد، أحتّ هيئة الأمم للعمل معاً لتبني معاهدة دوليّة ملزمة، تنظّم تطوير واستخدام الذّكاء الاصطناعيّ بأشكاله المتعدّدة. وبطبيعة الحال، ينبغي ألاّ يكون هدف التّنظيم منع الممارسات السيّئة فحسب، بل ينبغي أيضاً أن يكون تشجيع الممارسات الجيدة، وتحفيز الأساليب الجديدة والإبداعيّة، وتسهيل المبادرات الشّخصيّة

وفي النهاية، في البحث عن النماذج التنظيمية التي يمكن أن توفر توجيهًا أخلاقيًا لمطوري التكنولوجيات الرقمية، من الضروري تحديد القيم الإنسانية التي يجب أن تكون على أساس التزام المجتمعات لصياغة وتبني وتطبيق الأطر التشريعية اللازمة. العمل على صياغة مبادئ توجيهية أخلاقية لإنتاج أشكال الذكاء الاصطناعي لا يمكن أن يتجاهل النظر في أسئلة أعمق تتعلق بمعنى الحياة الإنسانية، وحماية حقوق الإنسان الأساسية، والسعي لتحقيق العدل والسلام. عملية التمييز الأخلاقي والقانوني هذه يمكن أن تكون فرصة ثمينة للتفكير المشترك في الدور الذي يجب أن تلعبه التكنولوجيا في حياتنا الفردية والجماعية، وكيف يمكن أن يؤدي استخدامها إلى المساهمة في خلق عالم فيه مزيد من المساواة والإنسانية. ولهذا السبب، في المناقشات في تنظيم الذكاء الاصطناعي، يجب الأخذ بعين الاعتبار أصوات جميع الأطراف المعنية، بما في ذلك الفقراء والمهمشين وغيرهم ممن يظلون غالبًا لا صوت لهم في عمليات صنع القرارات العالمية.

\* \* \* \* \*

أتمنى أن يشجعنا هذا التفكير حتى يكون التقدم في تطوير أشكال الذكاء الاصطناعي خادمًا في نهاية الأمر لقضية الأخوة الإنسانية والسلام. إنها ليست مسؤولية بعض الناس القليلين، بل مسؤولية الأسرة البشرية بأكملها. السلام، في الواقع، هو ثمرة العلاقات التي تعترف بالآخرين وتقبلهم في كرامتهم غير القابلة للمساومة، وهو ثمرة التعاون والالتزام في السعي لتحقيق التنمية المتكاملة لجميع الأشخاص وجميع الشعوب.

صلاحي في بداية السنة الجديدة هي أن التطور السريع لأشكال الذكاء الاصطناعي لا يزيد من عدم المساواة والمظالم الكبيرة الموجودة أصلًا في العالم، بل يساهم في وضع حد للحروب والصراعات، ويخفف من أشكال الآلام الكثيرة التي تتبلى الأسرة البشرية. أتمنى أن يتعاون المؤمنون المسيحيون، والمؤمنون من مختلف الديانات، والرجال والنساء ذوو الإرادة الصالحة، كلهم منسجمين معًا، ليغتنموا الفرص ويواجهوا التحديات التي تثيرها الثورة الرقمية، وليسلموا الأجيال القادمة عالمًا فيه مزيد من التضامن والعدل والسلام.

من حاضرة الفاتيكان، يوم 8 كانون الأول/ديسمبر من عام 2023.

فرنسيس

[1] رقم 33.

[2] المرجع نفسه، رقم 57.

[3] رسالة عامة بابوية، كُنْ مُسَبِّحًا (24 أيار/مايو 2015)، 104.

[4] راجع المرجع نفسه، 114.

[5] لقاء مع المشاركين في لقاء "27" (Minerva Dialogues) آذار/مارس 2023.

[6] راجع المرجع نفسه.

[7] راجع رسالة إلى الرئيس التنفيذي "للمنتدى الاقتصادي العالمي" في دافوس-كلوسترز (12 كانون الثاني/يناير 2018).

[8] راجع رسالة عامة بابوية، كُنْ مُسَبِّحًا، 194؛ كلمة للمشاركين في ندوة "الخير العام في العصر الرقمي" (27 أيلول/سبتمبر 2019).

[9] الإرشاد الرسولي، فرح الإنجيل (24 تشرين الثاني/نوفمبر 2013)، 233.

[10] راجع رسالة عامة بابوية، كُنْ مُسَبِّحًا، 54.

[11] راجع كلمة للمشاركين في مؤتمر الأكاديمية البابوية للحياة (28 شباط/فبراير 2020).

[12] راجع المرجع نفسه.

[13] رسالة عامة بابوية، كَلْنَا إِخْوَةَ - 3 Fratelli tutti تشرين الأول/أكتوبر 2020، 27.

[14] راجع المرجع نفسه.

[15] راجع المرجع نفسه، 170-175.

[16] راجع رسالة عامة بابوية، كُنْ مُسَبِّحًا، 177.

[01927-AR.01] [Testo originale: Italiano]

[B0884-XX.02]