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. 0504

Venerdì 14.06.2024

Sommario:

♦ Partecipazione del Santo Padre Francesco al G7 a Borgo Egnazia

♦ Partecipazione del Santo Padre Francesco al G7 a Borgo Egnazia

Versione integrale del Discorso preparato dal Santo Padre

Discorso pronunciato dal Santo Padre al G7

Questa mattina, lasciata Casa Santa Marta, il Santo Padre Francesco è partito in elicottero alle ore 10.30 dall'eliporto vaticano per partecipare al G7 a Borgo Egnazia, in Puglia.

Al Suo arrivo al campo sportivo, alle ore 12.10, il Santo Padre è stato accolto dal Presidente del Consiglio dei Ministri della Repubblica Italiana, On.le Giorgia Meloni.

Successivamente Papa Francesco si è trasferito in golf car alla residenza riservata dove erano previsti colloqui bilaterali con la Sig.ra Kristalina Georgieva, Direttore Generale dell'International Monetary Fund, con il Presidente della Repubblica di Ucraina, S.E. il Sig. Volodymyr Zelenskyy, con il Presidente della Repubblica Francese, S.E. il Sig. Emmanuel Macron, e con il Primo Ministro del Canada, S.E. il Sig. Justin Trudeau.

L'On.le Giorgia Meloni ha accolto ufficialmente il Santo Padre nella corte di Borgo Egnazia e, alle ore 14.55, nella Sala Arena, il Papa ha aperto la Sessione comune dei lavori del G7 e ha pronunciato il Suo discorso a cui hanno fatto seguito gli altri interventi.

Al termine sono proseguiti gli incontri bilaterali: a margine della Sessione pomeridiana appena conclusa, il Santo Padre ha incontrato il Primo Ministro dell'India, S.E. il Sig. Narendra Modi, e il Presidente della Repubblica di Turchia, S.E. il Sig. Recep Tayyp Erdoğan; e, successivamente, dopo la foto ufficiale di tutti i Partecipanti, il

Presidente del Kenya, S.E. il Sig. William Samoei Ruto; il Presidente del Brasile, S.E. il Sig. Luiz Inácio Lula da Silva; e il Presidente degli Stati Uniti d'America, S.E. il Sig. Joseph Biden.

A conclusione degli incontri bilaterali, alle ore 20.43, il Santo Padre è partito in elicottero dal campo sportivo di Borgo Egnazia per far rientro in Vaticano.

Pubblichiamo di seguito la versione integrale del discorso che il Santo Padre Francesco ha preparato per l'occasione e il discorso pronunciato nel corso della Sessione comune dei lavori del G7:

Versione integrale del Discorso preparato dal Santo Padre

- Testo in lingua italiana
- 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

Testo in lingua italiana

Uno strumento affascinante e tremendo

Gentili Signore, illustri Signori!

Mi rivolgo oggi a Voi, Leader del Forum Intergovernativo del G7, con una riflessione sugli effetti dell'intelligenza artificiale sul futuro dell'umanità.

«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)»[1]. La scienza e la tecnologia sono dunque prodotti straordinari del potenziale creativo di noi esseri umani[2].

Ebbene, è proprio dall'utilizzo di questo potenziale creativo che Dio ci ha donato che viene alla luce l'intelligenza artificiale.

Quest'ultima, come è noto, è uno strumento estremamente potente, impiegato in tantissime aree dell'agire umano: dalla medicina al mondo del lavoro, dalla cultura all'ambito della comunicazione, dall'educazione alla politica. Ed è ora lecito ipotizzare che il suo uso influenzerà sempre di più il nostro modo di vivere, le nostre relazioni sociali e nel futuro persino la maniera in cui concepiamo la nostra identità di esseri umani[3].

Il tema dell'intelligenza artificiale è, tuttavia, spesso percepito come ambivalente: da un lato, entusiasma per le possibilità che offre, dall'altro genera timore per le conseguenze che lascia presagire. A questo proposito si può dire che tutti noi siamo, anche se in misura diversa, attraversati da due emozioni: siamo entusiasti, quando

immaginiamo i progressi che dall'intelligenza artificiale possono derivare, ma, al tempo stesso, siamo impauriti quando constatiamo i pericoli inerenti al suo uso[4].

Non possiamo, del resto, dubitare che l'avvento dell'intelligenza artificiale rappresenti una vera e propria rivoluzione cognitivo-industriale, che contribuirà alla creazione di un nuovo sistema sociale caratterizzato da complesse trasformazioni epocali. Ad esempio, l'intelligenza artificiale potrebbe permettere una democratizzazione dell'accesso al sapere, il progresso esponenziale della ricerca scientifica, la possibilità di delegare alle macchine i lavori usuranti; ma, al tempo stesso, essa potrebbe portare con sé una più grande ingiustizia fra nazioni avanzate e nazioni in via di sviluppo, fra ceti sociali dominanti e ceti sociali oppressi, mettendo così in pericolo la possibilità di una "cultura dell'incontro" a vantaggio di una "cultura dello scarto".

La portata di queste complesse trasformazioni è ovviamente legata al rapido sviluppo tecnologico dell'intelligenza artificiale stessa.

Proprio questo vigoroso avanzamento tecnologico rende l'intelligenza artificiale *uno strumento affascinante* e *tremendo* al tempo stesso ed impone una riflessione all'altezza della situazione.

In tale direzione forse si potrebbe partire dalla costatazione che l'intelligenza artificiale è innanzitutto *uno strumento*. E viene spontaneo affermare che i benefici o i danni che essa porterà dipenderanno dal suo impiego.

Questo è sicuramente vero, poiché così è stato per ogni utensile costruito dall'essere umano sin dalla notte dei tempi.

Questa nostra capacità di costruire utensili, in una quantità e complessità che non ha pari tra i viventi, fa parlare di una *condizione tecno-umana*: l'essere umano ha da sempre mantenuto una relazione con l'ambiente mediata dagli strumenti che via via produceva. Non è possibile separare la storia dell'uomo e della civilizzazione dalla storia di tali strumenti. Qualcuno ha voluto leggere in tutto ciò una sorta di mancanza, un deficit, dell'essere umano, come se, a causa di tale carenza, fosse costretto a dare vita alla tecnologia[5]. Uno sguardo attento e oggettivo in realtà ci mostra l'opposto. Viviamo una condizione di ulteriorità rispetto al nostro essere biologico; siamo esseri sbilanciati verso il fuori-di-noi, anzi radicalmente aperti all'oltre. Da qui prende origine la nostra apertura agli altri e a Dio; da qui nasce il potenziale creativo della nostra intelligenza in termini di cultura e di bellezza; da qui, da ultimo, si origina la nostra capacità tecnica. La tecnologia è così una traccia di questa nostra ulteriorità.

Tuttavia, l'uso dei nostri utensili non sempre è univocamente rivolto al bene. Anche se l'essere umano sente dentro di sé una vocazione all'oltre e alla conoscenza vissuta come strumento di bene al servizio dei fratelli e delle sorelle e della *casa comune* (cfr *Gaudium et spes*, 16), non sempre questo accade. Anzi, non di rado, proprio grazie alla sua radicale libertà, l'umanità ha pervertito i fini del suo essere trasformandosi in nemica di sé stessa e del pianeta[6]. Stessa sorte possono avere gli strumenti tecnologici. Solo se sarà garantita la loro vocazione al servizio dell'umano, gli strumenti tecnologici riveleranno non solo la grandezza e la dignità unica dell'essere umano, ma anche il mandato che quest'ultimo ha ricevuto di "coltivare e custodire" (cfr *Gen* 2,15) il pianeta e tutti i suoi abitanti. Parlare di tecnologia è parlare di cosa significhi essere umani e quindi di quella nostra unica condizione tra libertà e responsabilità, cioè vuol dire parlare di etica.

Quando i nostri antenati, infatti, affilarono delle pietre di selce per costruire dei coltelli, li usarono sia per tagliare il pellame per i vestiti sia per uccidersi gli uni gli altri. Lo stesso si potrebbe dire di altre tecnologie molto più avanzate, quali l'energia prodotta dalla fusione degli atomi come avviene sul Sole, che potrebbe essere utilizzata certamente per produrre energia pulita e rinnovabile ma anche per ridurre il nostro pianeta in un cumulo di cenere.

L'intelligenza artificiale, però, è uno strumento ancora più complesso. Direi quasi che si tratta di uno strumento *sui generis*. Così, mentre l'uso di un utensile semplice (come il coltello) è sotto il controllo dell'essere umano che lo utilizza e solo da quest'ultimo dipende un suo buon uso, l'intelligenza artificiale, invece, può adattarsi autonomamente al compito che le viene assegnato e, se progettata con questa modalità, operare scelte

indipendenti dall'essere umano per raggiungere l'obiettivo prefissato[7].

Conviene sempre ricordare che la macchina può, in alcune forme e con questi nuovi mezzi, produrre delle scelte algoritmiche. Ciò che la macchina fa è una scelta tecnica tra più possibilità e si basa o su criteri ben definiti o su inferenze statistiche. L'essere umano, invece, non solo sceglie, ma in cuor suo è capace di decidere. La decisione è un elemento che potremmo definire maggiormente strategico di una scelta e richiede una valutazione pratica. A volte, spesso nel difficile compito del governare, siamo chiamati a decidere con conseguenze anche su molte persone. Da sempre la riflessione umana parla a tale proposito di saggezza, la *phronesis* della filosofia greca e almeno in parte la sapienza della Sacra Scrittura. Di fronte ai prodigi delle macchine, che sembrano saper scegliere in maniera indipendente, dobbiamo aver ben chiaro che all'essere umano deve sempre rimanere la decisione, anche con i toni drammatici e urgenti con cui a volte questa si presenta nella nostra vita. Condanneremmo l'umanità a un futuro senza speranza, se sottraessimo alle persone la capacità di decidere su loro stesse e sulla loro vita condannandole a dipendere dalle scelte delle macchine. Abbiamo bisogno di garantire e tutelare uno spazio di controllo significativo dell'essere umano sul processo di scelta dei programmi di intelligenza artificiale: ne va della stessa dignità umana.

Proprio su questo tema permettetemi di insistere: in un dramma come quello dei conflitti armati è urgente ripensare lo sviluppo e l'utilizzo di dispositivi come le cosiddette "armi letali autonome" per bandirne l'uso, cominciando già da un impegno fattivo e concreto per introdurre un sempre maggiore e significativo controllo umano. Nessuna macchina dovrebbe mai scegliere se togliere la vita ad un essere umano.

C'è da aggiungere, inoltre, che il buon uso, almeno delle forme avanzate di intelligenza artificiale, non sarà pienamente sotto il controllo né degli utilizzatori né dei programmatori che ne hanno definito gli scopi originari al momento dell'ideazione. E questo è tanto più vero quanto è altamente probabile che, in un futuro non lontano, i programmi di intelligenze artificiali potranno comunicare direttamente gli uni con gli altri, per migliorare le loro *performance*. E, se in passato, gli esseri umani che hanno modellato utensili semplici hanno visto la loro esistenza modellata da questi ultimi – il coltello ha permesso loro di sopravvivere al freddo ma anche di sviluppare l'arte della guerra – adesso che gli esseri umani hanno modellato uno strumento complesso vedranno quest'ultimo modellare ancora di più la loro esistenza[8].

Il meccanismo basilare dell'intelligenza artificiale

Vorrei ora soffermarmi brevemente sulla complessità dell'intelligenza artificiale. Nella sua essenza l'intelligenza artificiale è un utensile disegnato per la risoluzione di un problema e funziona per mezzo di un concatenamento logico di operazioni algebriche, effettuato su categorie di dati, che sono raffrontati per scoprire delle correlazioni, migliorandone il valore statistico, grazie a un processo di auto-apprendimento, basato sulla ricerca di ulteriori dati e sull'auto-modifica delle sue procedure di calcolo.

L'intelligenza artificiale è così disegnata per risolvere dei problemi specifici, ma per coloro che la utilizzano è spesso irresistibile la tentazione di trarre, a partire dalle soluzioni puntuali che essa propone, delle deduzioni generali, persino di ordine antropologico.

Un buon esempio è l'uso dei programmi disegnati per aiutare i magistrati nelle decisioni relative alla concessione dei domiciliari a detenuti che stanno scontando una pena in un istituto carcerario. In questo caso, si chiede all'intelligenza artificiale di prevedere la probabilità di recidiva del crimine commesso da parte di un condannato a partire da categorie prefissate (tipo di reato, comportamento in prigione, valutazione psicologiche ed altro), permettendo all'intelligenza artificiale di avere accesso a categorie di dati inerenti alla vita privata del detenuto (origine etnica, livello educativo, linea di credito ed altro). L'uso di una tale metodologia – che rischia a volte di delegare *de facto* a una macchina l'ultima parola sul destino di una persona – può portare con sé implicitamente il riferimento ai pregiudizi insiti alle categorie di dati utilizzati dall'intelligenza artificiale.

L'essere classificato in un certo gruppo etnico o, più prosaicamente, l'aver commesso anni prima un'infrazione minore (il non avere pagato, per esempio, una multa per una sosta vietata), influenzerà, infatti, la decisione circa la concessione dei domiciliari. Al contrario, l'essere umano è sempre in evoluzione ed è capace di sorprendere

con le sue azioni, cosa di cui la macchina non può tenere conto.

C'è da far presente poi che applicazioni simili a questa appena citata subiranno un'accelerazione grazie al fatto che i programmi di intelligenza artificiale saranno sempre più dotati della capacità di interagire direttamente con gli esseri umani (*chatbots*), sostenendo conversazioni con loro e stabilendo rapporti di vicinanza con loro, spesso molto piacevoli e rassicuranti, in quanto tali programmi di intelligenza artificiale saranno disegnati per imparare a rispondere, in forma personalizzata, ai bisogni fisici e psicologici degli esseri umani.

Dimenticare che l'intelligenza artificiale non è un altro essere umano e che essa non può proporre principi generali, è spesso un grave errore che trae origine o dalla profonda necessità degli esseri umani di trovare una forma stabile di compagnia o da un loro presupposto subcosciente, ossia dal presupposto che le osservazioni ottenute mediante un meccanismo di calcolo siano dotate delle qualità di certezza indiscutibile e di universalità indubbia.

Questo presupposto, tuttavia, è azzardato, come dimostra l'esame dei limiti intrinseci del calcolo stesso. L'intelligenza artificiale usa delle operazioni algebriche da effettuarsi secondo una sequenza logica (per esempio, se il valore di X è superiore a quello di Y, moltiplica X per Y; altrimenti dividi X per Y). Questo metodo di calcolo – il cosiddetto "algoritmo" – non è dotato né di oggettività né di neutralità[9]. Essendo infatti basato sull'algebra, può esaminare solo realtà formalizzate in termini numerici[10].

Non va dimenticato, inoltre, che gli algoritmi disegnati per risolvere problemi molto complessi sono così sofisticati da rendere arduo agli stessi programmatori la comprensione esatta del come essi riescano a raggiungere i loro risultati. Questa tendenza alla sofisticazione rischia di accelerarsi notevolmente con l'introduzione di computer quantistici che non opereranno con circuiti binari (semiconduttori o microchip), ma secondo le leggi, alquanto articolate, della fisica quantistica.

D'altronde, la continua introduzione di microchip sempre più performanti è diventata già una delle cause del predominio dell'uso dell'intelligenza artificiale da parte delle poche nazioni che ne sono dotate.

Sofisticate o meno che siano, la qualità delle risposte che i programmi di intelligenza artificiale forniscono dipendono in ultima istanza dai dati che essi usano e come da questi ultimi vengono strutturati.

Mi permetto di segnalare, infine, un ultimo ambito in cui emerge chiaramente la complessità del meccanismo della cosiddetta intelligenza artificiale generativa (*Generative Artificial Intelligence*). Nessuno dubita che oggi sono a disposizione magnifici strumenti di accesso alla conoscenza che permettono persino il *self-learning* e il *self-tutoring* in una miriade di campi. Molti di noi sono rimasti colpiti dalle applicazioni facilmente disponibili on-line per comporre un testo o produrre un'immagine su qualsiasi tema o soggetto. Particolarmente attratti da questa prospettiva sono gli studenti che, quando devono preparare degli elaborati, ne fanno un uso sproporzionato.

Questi alunni, che spesso sono molto più preparati e abituati all'uso dell'intelligenza artificiale dei loro professori, dimenticano, tuttavia, che la cosiddetta intelligenza artificiale generativa, in senso stretto, non è propriamente "generativa". Quest'ultima, in verità, cerca nei *big data* delle informazioni e le confeziona nello stile che le è stato richiesto. Non sviluppa concetti o analisi nuove. Ripete quelle che trova, dando loro una forma accattivante. E più trova ripetuta una nozione o una ipotesi, più la considera legittima e valida. Più che "generativa", essa è quindi "rafforzativa", nel senso che riordina i contenuti esistenti, contribuendo a consolidarli, spesso senza controllare se contengano errori o preconcetti.

In questo modo, non solo si corre il rischio di legittimare delle *fake news* e di irrobustire il vantaggio di una cultura dominante, ma di minare altresì il processo educativo *in nuce*. L'educazione che dovrebbe fornire agli studenti la possibilità di una riflessione autentica rischia di ridursi a una ripetizione di nozioni, che verranno sempre di più valutate come inoppugnabili, semplicemente in ragione della loro continua riproposizione[11].

Rimettere al centro la dignità della persona in vista di una proposta etica condivisa

A quanto già detto va ora aggiunta un'osservazione più generale. La stagione di innovazione tecnologica che stiamo attraversando, infatti, si accompagna a una particolare e inedita congiuntura sociale: sui grandi temi del vivere sociale si riesce con sempre minore facilità a trovare intese. Anche in comunità caratterizzate da una certa continuità culturale, si creano spesso accesi dibattiti e confronti che rendono difficile produrre riflessioni e soluzioni politiche condivise, volte a cercare ciò che è bene e giusto. Oltre la complessità di legittime visioni che caratterizzano la famiglia umana, emerge un fattore che sembra accomunare queste diverse istanze. Si registra come uno smarrimento o quantomeno un'eclissi del senso dell'umano e un'apparente insignificanza del concetto di dignità umana[12]. Sembra che si stia perdendo il valore e il profondo significato di una delle categorie fondamentali dell'Occidente: la categoria di persona umana. Ed è così che in questa stagione in cui i programmi di intelligenza artificiale interrogano l'essere umano e il suo agire, proprio la debolezza dell'*ethos* connesso alla percezione del valore e della dignità della persona umana rischia di essere il più grande *vulnus* nell'implementazione e nello sviluppo di questi sistemi. Non dobbiamo dimenticare infatti che nessuna innovazione è neutrale. La tecnologia nasce per uno scopo e, nel suo impatto con la società umana, rappresenta sempre una forma di ordine nelle relazioni sociali e una disposizione di potere, che abilita qualcuno a compiere azioni e impedisce ad altri di compierne altre. Questa costitutiva dimensione di potere della tecnologia include sempre, in una maniera più o meno esplicita, la visione del mondo di chi l'ha realizzata e sviluppata.

Questo vale anche per i programmi di intelligenza artificiale. Affinché questi ultimi siano strumenti per la costruzione del bene e di un domani migliore, debbono essere sempre ordinati al bene di ogni essere umano. Devono avere un'ispirazione etica.

La decisione etica, infatti, è quella che tiene conto non solo degli esiti di un'azione, ma anche dei valori in gioco e dei doveri che da questi valori derivano. Per questo ho salutato con favore la firma a Roma, nel 2020, della *Rome Call for AI Ethics*[13] e il suo sostegno a quella forma di moderazione etica degli algoritmi e dei programmi di intelligenza artificiale che ho chiamato "algoretica" [14]. In un contesto plurale e globale, in cui si mostrano anche sensibilità diverse e gerarchie plurali nelle scale dei valori, sembrerebbe difficile trovare un'unica gerarchia di valori. Ma nell'analisi etica possiamo ricorrere anche ad altri tipi di strumenti: se facciamo fatica a definire un solo insieme di valori globali, possiamo però trovare dei principi condivisi con cui affrontare e sciogliere eventuali dilemmi o conflitti del vivere.

Per questa ragione è nata la *Rome Call:* nel termine "algoretica" si condensano una serie di principi che si dimostrano essere una piattaforma globale e plurale in grado di trovare il supporto di culture, religioni, organizzazioni internazionali e grandi aziende protagoniste di questo sviluppo.

La politica di cui c'è bisogno

Non possiamo, quindi, nascondere il rischio concreto, poiché insito nel suo meccanismo fondamentale, che l'intelligenza artificiale limiti la visione del mondo a realtà esprimibili in numeri e racchiuse in categorie preconfezionate, estromettendo l'apporto di altre forme di verità e imponendo modelli antropologici, socio-economici e culturali uniformi. Il paradigma tecnologico incarnato dall'intelligenza artificiale rischia allora di fare spazio a un paradigma ben più pericoloso, che ho già identificato con il nome di "paradigma tecnocratico"[15]. Non possiamo permettere a uno strumento così potente e così indispensabile come l'intelligenza artificiale di rinforzare un tale paradigma, ma anzi, dobbiamo fare dell'intelligenza artificiale un baluardo proprio contro la sua espansione.

Ed è proprio qui che è urgente l'azione politica, come ricorda l'Enciclica *Fratelli tutti*. Certamente «per molti la politica oggi è una brutta parola, e non si può ignorare che dietro questo fatto ci sono spesso gli errori, la corruzione, l'inefficienza di alcuni politici. A ciò si aggiungono le strategie che mirano a indebolirla, a sostituirla con l'economia o a dominarla con qualche ideologia. E tuttavia, può funzionare il mondo senza politica? Può trovare una via efficace verso la fraternità universale e la pace sociale senza una buona politica?»[16].

La nostra risposta a queste ultime domande è: no! La politica serve! Voglio ribadire in questa occasione che «davanti a tante forme di politica meschine e tese all'interesse immediato […] la grandezza politica si mostra quando, in momenti difficili, si opera sulla base di grandi principi e pensando al bene comune a lungo termine. Il potere politico fa molta fatica ad accogliere questo dovere in un progetto di Nazione e ancora di più in un progetto comune per l'umanità presente e futura»[17].

Gentili Signore, illustri Signori!

Questa mia riflessione sugli effetti dell'intelligenza artificiale sul futuro dell'umanità ci conduce così alla considerazione dell'importanza della "sana politica" per guardare con speranza e fiducia al nostro avvenire. Come ho già detto altrove, «la società mondiale ha gravi carenze strutturali che non si risolvono con rattoppi o soluzioni veloci meramente occasionali. Ci sono cose che devono essere cambiate con reimpostazioni di fondo e trasformazioni importanti. Solo una sana politica potrebbe averne la guida, coinvolgendo i più diversi settori e i più vari saperi. In tal modo, un'economia integrata in un progetto politico, sociale, culturale e popolare che tenda al bene comune può "aprire la strada a opportunità differenti, che non implicano di fermare la creatività umana e il suo sogno di progresso, ma piuttosto di incanalare tale energia in modo nuovo" (*Laudato si'*, 191)» [18].

Questo è proprio il caso dell'intelligenza artificiale. Spetta ad ognuno farne buon uso e spetta alla politica creare le condizioni perché un tale buon uso sia possibile e fruttuoso.

Grazie.

---

[1] *Messaggio per la LVII Giornata Mondiale della Pace del 1° gennaio 2024*, 1.

[2] Cfr *ibid*.

[3] Cfr *ivi*, 2.

[4] Questa ambivalenza fu già scorta da Papa San Paolo VI nel suo *Discorso al personale del "Centro Automazione Analisi Linguistica" dell'Aloysianum*, del 19 giugno 1964.

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

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

[7] Cfr *Messaggio per la LVII Giornata Mondiale della Pace del 1° gennaio 2024*, 3.

[8] Le intuizioni di Marshall McLuhan e di John M. Culkin sono particolarmente pertinenti alle conseguenze dell'uso dell'intelligenza artificiale.

[9] Cfr *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[10] Cfr *Messaggio per la LVII Giornata Mondiale della Pace del 1° gennaio 2024*, 4.

[11] Cfr *ivi*, 3 e 7.

[12] Cfr Dicastero per la Dottrina della Fede, Dichiarazione *Dignitas infinita* circa la dignità umana (2 aprile 2024).

[13] Cfr *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[14] Cfr *Discorso ai partecipanti al Convegno "Promoting Digital Child Dignity – From Concet to Action"*, 14 novembre 2019; *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[15] Per una più ampia esposizione, rimando alla mia Lettera Enciclica *Laudato si'* sulla cura della casa comune del 24 maggio 2015.

[16] Lettera enc. *Fratelli tutti* sulla fraternità e l'amicizia sociale (3 ottobre 2020), 176.

[17] *Ivi*, 178.

[18] *Ivi*, 179.

[01030-IT.01] [Testo originale: Italiano]

Traduzione in lingua francese

Un outil fascinant et redoutable

Mesdames et Messieurs !

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

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

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

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

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

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

“culture du rejet”.

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

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

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

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

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

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

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

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

Il faut toujours garder à l’esprit que la machine peut, sous certaines formes et par ces nouveaux moyens, produire des choix algorithmiques. Ce que fait la machine est un choix technique entre plusieurs possibilités et se base soit sur des critères bien définis, soit sur des déductions statistiques. L’être humain, quant à lui, non seulement choisit, mais dans son cœur il est capable de décider. La décision est un élément que nous pourrions concevoir plus stratégique qu’un choix et nécessite une évaluation pratique. Parfois, et bien souvent dans la tâche difficile de gouverner, nous sommes appelés à prendre des décisions qui ont des conséquences pour de nombreuses personnes. Depuis toujours la réflexion humaine parle à ce propos de sagesse, la *phronesis* de la

philosophie grecque et au moins en partie la sagesse de l'Écriture Sainte. Face aux prodiges des machines, qui semblent capables de choisir de manière autonome, nous devons être clairs sur le fait que la décision doit toujours être laissée à l'être humain, même dans une tournure dramatique et urgente avec laquelle elle se présente parfois dans nos vies. Nous condamnerions l'humanité à un avenir sans espoir si nous retirions aux gens la capacité de décider d'eux-mêmes et de leur vie, les condamnant à dépendre des choix des machines. Nous devons garantir et protéger un espace de contrôle humain significatif sur le processus de choix des programmes d'intelligence artificielle : la dignité humaine elle-même en dépend.

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

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

Le mécanisme de base de l'intelligence artificielle

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

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

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

Le fait d'être classé dans un certain groupe ethnique ou, plus prosaïquement, d'avoir commis un délit mineur des années auparavant (ne pas avoir payé, par exemple, une amende pour un stationnement interdit) influencera, en effet, la décision concernant le fait de procéder ou non à une assignation à résidence. Au contraire, l'être humain évolue en permanence et se montre capable de surprendre par ses actes, chose que la machine ne peut pas prendre en compte.

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

d'intelligence artificielle seront conçus pour apprendre à répondre, de manière personnalisée, aux besoins physiques et psychologiques des êtres humains.

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

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

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

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

Enfin, je voudrais souligner un dernier domaine dans lequel apparaît clairement la complexité du mécanisme de l'intelligence artificielle dite générative (*Generative Artificial Intelligence*). Nul ne doute qu'il existe aujourd'hui de magnifiques outils d'accès à la connaissance qui permettent même l'auto-apprentissage et l'auto-tutorat dans une myriade de domaines. Beaucoup d'entre nous ont été impressionnés par les applications facilement disponibles en ligne pour composer un texte ou produire une image sur n'importe quel thème ou sujet. Les étudiants se montrent particulièrement attirés par cette perspective qui, lorsqu'ils doivent préparer des travaux, en font un usage disproportionné.

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

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

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

Il convient maintenant d'ajouter une observation plus générale à ce qui a déjà été dit. La saison d'innovation technologique que nous traversons actuellement s'accompagne en effet d'une conjoncture sociale particulière et sans précédent : sur les grandes questions de la vie sociale, il est de plus en plus difficile de trouver des accords. Même dans les communautés caractérisées par une certaine continuité culturelle, des débats

passionnés et des confrontations surgissent souvent, rendant difficile la production de réflexions et de solutions politiques partagées visant à rechercher ce qui est bon et juste. Au-delà de la complexité des visions légitimes qui caractérisent la famille humaine, un facteur émerge qui semble unir ces différentes instances. Nous assistons à une disparition ou du moins à une éclipse du sens de l'humain et à une apparente insignifiance du concept de dignité humaine[12]. Il semble que nous perdons la valeur et le sens profond de l'une des catégories fondamentales de l'Occident : la catégorie de la personne humaine. En ce moment où les programmes d'intelligence artificielle remettent en question l'être humain et son agir, c'est précisément la faiblesse de l'*ethos* lié à la perception de la valeur et de la dignité de la personne humaine qui risque d'être le plus grand *vulnus* dans la mise en œuvre et le développement de ces systèmes. En effet, il ne faut pas oublier qu'aucune innovation n'est neutre. La technologie naît dans un but précis et, dans son impact sur la société humaine, elle représente toujours une forme d'ordre dans les relations sociales et une disposition de pouvoir, permettant à certains d'accomplir des actions et empêchant d'autres d'en accomplir d'autres. Cette dimension constitutive de pouvoir de la technologie comprend toujours, de manière plus ou moins explicite, la vision du monde de ceux qui l'ont conçue et développée.

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

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

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

sont des acteurs de ce développement.

La politique dont nous avons besoin

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

Et c'est précisément là que l'action politique est urgente, comme le rappelle l'encyclique *Fratelli tutti*. Certes, « pour beaucoup de personnes, la politique est aujourd'hui un vilain mot et on ne peut pas ignorer qu'à la base de ce fait, il y a souvent les erreurs, la corruption, l'inefficacité de certains hommes politiques. À cela s'ajoutent les stratégies qui cherchent à affaiblir la politique, à la remplacer par l'économie ou la soumettre à quelque idéologie. Mais le monde peut-il fonctionner sans la politique ? Peut-il y avoir un chemin approprié vers la fraternité universelle et la paix sociale sans une bonne politique ? »[16].

Notre réponse à ces dernières questions est : non ! La politique est nécessaire ! Je tiens à réaffirmer à cette

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

Mesdames, Messieurs !

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

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

Merci.

---

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

[2] Cf. *Ibid*.

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

[4] Cette ambivalence avait déjà été soulignée par le Pape saint Paul VI dans son *Discours au personnel du "Centre d'automatisation des analyses linguistiques" de l'Aloysianum*, du 19 juin 1964.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[01030-FR.01] [Texte original: Italien]

Traduzione in lingua inglese

An exciting and fearsome tool

Esteemed ladies and gentlemen,

I address you today, the leaders of the Intergovernmental Forum of the G7, concerning the effects of artificial intelligence on the future of humanity.

"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)".[1] Science and technology are therefore brilliant products of the creative potential of human beings.[2]

Indeed, artificial intelligence arises precisely from the use of this God-given creative potential.

As we know, artificial intelligence is an extremely powerful tool, employed in many kinds of human activity: from medicine to the world of work; from culture to the field of communications; from education to politics. It is now safe to assume that its use will increasingly influence the way we live, our social relationships and even the way we conceive of our identity as human beings.[3]

The question of artificial intelligence, however, is often perceived as ambiguous: on the one hand, it generates excitement for the possibilities it offers, while on the other it gives rise to fear for the consequences it foreshadows. In this regard, we could say that all of us, albeit to varying degrees, experience two emotions: we are enthusiastic when we imagine the advances that can result from artificial intelligence but, at the same time, we are fearful when we acknowledge the dangers inherent in its use.[4]

After all, we cannot doubt that the advent of artificial intelligence represents a true cognitive-industrial revolution, which will contribute to the creation of a new social system characterised by complex epochal transformations. For example, artificial intelligence could enable a democratization of access to knowledge, the exponential advancement of scientific research and the possibility of giving demanding and arduous work to machines. Yet at the same time, it could bring with it a greater injustice between advanced and developing nations or between dominant and oppressed social classes, raising the dangerous possibility that a "throwaway culture" be preferred to a "culture of encounter".

The significance of these complex transformations is clearly linked to the rapid technological development of

artificial intelligence itself.

It is precisely this powerful technological progress that makes artificial intelligence at the same time *an exciting and fearsome tool*, and demands a reflection that is up to the challenge it presents.

In this regard, perhaps we could start from the observation that artificial intelligence is above all else *a tool*. And it goes without saying that the benefits or harm it will bring will depend on its use.

This is surely the case, for it has been this way with every tool fashioned by human beings since the dawn of time.

Our ability to fashion tools, in a quantity and complexity that is unparalleled among living things, speaks of a *techno-human condition*: human beings have always maintained a relationship with the environment mediated by the tools they gradually produced. It is not possible to separate the history of men and women and of civilization from the history of these tools. Some have wanted to read into this a kind of shortcoming, a deficit, within human beings, as if, because of this deficiency, they were forced to create technology.[5] A careful and objective view actually shows us the opposite. We experience a state of “outwardness” with respect to our biological being: we are beings inclined toward what lies outside-of-us, indeed we are radically open to the beyond. Our openness to others and to God originates from this reality, as does the creative potential of our intelligence with regard to culture and beauty. Ultimately, our technical capacity also stems from this fact. Technology, then, is a sign of our orientation towards the future.

The use of our tools, however, is not always directed solely to the good. Even if human beings feel within themselves a call to the beyond, and to knowledge as an instrument of good for the service of our brothers and sisters and our *common home* (cf. *Gaudium et Spes*, 16), this does not always happen. Due to its radical freedom, humanity has not infrequently corrupted the purposes of its being, turning into an enemy of itself and of the planet.[6] The same fate may befall technological tools. Only if their true purpose of serving humanity is ensured, will such tools reveal not only the unique grandeur and dignity of men and women, but also the command they have received to “till and keep” (cf. *Gen* 2:15) the planet and all its inhabitants. To speak of technology is to speak of what it means to be human and thus of our singular status as beings who possess both freedom and responsibility. This means speaking about ethics.

In fact, when our ancestors sharpened flint stones to make knives, they used them both to cut hides for clothing and to kill each other. The same could be said of other more advanced technologies, such as the energy produced by the fusion of atoms, as occurs within the Sun, which could be used to produce clean, renewable energy or to reduce our planet to a pile of ashes.

Artificial intelligence, however, is a still more complex tool. I would almost say that we are dealing with a tool *sui generis*. While the use of a simple tool (like a knife) is under the control of the person who uses it and its use for the good depends only on that person, artificial intelligence, on the other hand, can autonomously adapt to the task assigned to it and, if designed this way, can make choices independent of the person in order to achieve the intended goal.[7]

It should always be remembered that a machine can, in some ways and by these new methods, produce algorithmic choices. The machine makes a technical choice among several possibilities based either on well-defined criteria or on statistical inferences. Human beings, however, not only choose, but in their hearts are capable of deciding. A decision is what we might call a more strategic element of a choice and demands a practical evaluation. At times, frequently amid the difficult task of governing, we are called upon to make decisions that have consequences for many people. In this regard, human reflection has always spoken of wisdom, the *phronesis* of Greek philosophy and, at least in part, the wisdom of Sacred Scripture. Faced with the marvels of machines, which seem to know how to choose independently, we should be very clear that decision-making, even when we are confronted with its sometimes dramatic and urgent aspects, must always be left to the human person. We would condemn humanity to a future without hope if we took away people’s ability to make decisions about themselves and their lives, by dooming them to depend on the choices of machines. We

need to ensure and safeguard a space for proper human control over the choices made by artificial intelligence programs: human dignity itself depends on it.

Precisely in this regard, allow me to insist: in light of the tragedy that is armed conflict, it is urgent to reconsider the development and use of devices like the so-called “lethal autonomous weapons” and ultimately ban their use. This starts from an effective and concrete commitment to introduce ever greater and proper human control. No machine should ever choose to take the life of a human being.

It must be added, moreover, that the good use, at least of advanced forms of artificial intelligence, will not be fully under the control of either the users or the programmers who defined their original purposes at the time they were designed. This is all the more true because it is highly likely that, in the not-too-distant future, artificial intelligence programs will be able to communicate directly with each other to improve their performance. And if, in the past, men and women who fashioned simple tools saw their lives shaped by them – the knife enabled them to survive the cold but also to develop the art of warfare – now that human beings have fashioned complex tools they will see their lives shaped by them all the more.[8]

The basic mechanism of artificial intelligence

I would like now briefly to address the complexity of artificial intelligence. Essentially, artificial intelligence is a tool designed for problem solving. It works by means of a logical chaining of algebraic operations, carried out on categories of data. These are then compared in order to discover correlations, thereby improving their statistical value. This takes place thanks to a process of self-learning, based on the search for further data and the self-modification of its calculation processes.

Artificial intelligence is designed in this way in order to solve specific problems. Yet, for those who use it, there is often an irresistible temptation to draw general, or even anthropological, deductions from the specific solutions it offers.

An important example of this is the use of programs designed to help judges in deciding whether to grant home-confinement to inmates serving a prison sentence. In this case, artificial intelligence is asked to predict the likelihood of a prisoner committing the same crime(s) again. It does so based on predetermined categories (type of offence, behaviour in prison, psychological assessment, and others), thus allowing artificial intelligence to have access to categories of data relating to the prisoner’s private life (ethnic origin, educational attainment, credit rating, and others). The use of such a methodology – which sometimes risks *de facto* delegating to a machine the last word concerning a person’s future – may implicitly incorporate prejudices inherent in the categories of data used by artificial intelligence.

Being classified as part of a certain ethnic group, or simply having committed a minor offence years earlier (for example, not having paid a parking fine) will actually influence the decision as to whether or not to grant home-confinement. In reality, however, human beings are always developing, and are capable of surprising us by their actions. This is something that a machine cannot take into account.

It should also be noted that the use of applications similar to the one I have just mentioned will be used ever more frequently due to the fact that artificial intelligence programs will be increasingly equipped with the capacity to interact directly (*chatbots*) with human beings, holding conversations and establishing close relationships with them. These interactions may end up being, more often than not, pleasant and reassuring, since these artificial intelligence programs will be designed to learn to respond, in a personalised way, to the physical and psychological needs of human beings.

It is a frequent and serious mistake to forget that artificial intelligence is not another human being, and that it cannot propose general principles. This error stems either from the profound need of human beings to find a stable form of companionship, or from a subconscious assumption, namely the assumption that observations obtained by means of a calculating mechanism are endowed with the qualities of unquestionable certainty and unquestionable universality.

This assumption, however, is far-fetched, as can be seen by an examination of the inherent limitations of computation itself. Artificial intelligence uses algebraic operations that are carried out in a logical sequence (for example, if the value of X is greater than that of Y, multiply X by Y; otherwise divide X by Y). This method of calculation – the so-called “algorithm” – is neither objective nor neutral.[9] Moreover, since it is based on algebra, it can only examine realities formalised in numerical terms.[10]

Nor should it be forgotten that algorithms designed to solve highly complex problems are so sophisticated that it is difficult for programmers themselves to understand exactly how they arrive at their results. This tendency towards sophistication is likely to accelerate considerably with the introduction of quantum computers that will operate not with binary circuits (semiconductors or microchips) but according to the highly complex laws of quantum physics. Indeed, the continuous introduction of increasingly high-performance microchips has already become one of the reasons for the dominant use of artificial intelligence by those few nations equipped in this regard.

Whether sophisticated or not, the quality of the answers that artificial intelligence programs provide ultimately depends on the data they use and how they are structured.

Finally, I would like to indicate one last area in which the complexity of the mechanism of so-called Generative Artificial Intelligence clearly emerges. Today, no one doubts that there are magnificent tools available for accessing knowledge, which even allow for self-learning and self-tutoring in a myriad of fields. Many of us have been impressed by the easily available online applications for composing a text or producing an image on any theme or subject. Students are especially attracted to this, but make disproportionate use of it when they have to prepare papers.

Students are often much better prepared for, and more familiar with, using artificial intelligence than their teachers. Yet they forget that, strictly speaking, so-called generative artificial intelligence is not really “generative”. Instead, it searches big data for information and puts it together in the style required of it. It does not develop new analyses or concepts, but repeats those that it finds, giving them an appealing form. Then, the more it finds a repeated notion or hypothesis, the more it considers it legitimate and valid. Rather than being “generative”, then, it is instead “reinforcing” in the sense that it rearranges existing content, helping to consolidate it, often without checking whether it contains errors or preconceptions.

In this way, it not only runs the risk of legitimising fake news and strengthening a dominant culture’s advantage, but, in short, it also undermines the educational process itself. Education should provide students with the possibility of authentic reflection, yet it runs the risk of being reduced to a repetition of notions, which will increasingly be evaluated as unobjectionable, simply because of their constant repetition.[11]

Putting the dignity of the human person back at the centre, in light of a shared ethical proposal

A more general observation should now be added to what we have already said. The season of technological innovation in which we are currently living is accompanied by a particular and unprecedented social situation in which it is increasingly difficult to find agreement on the major issues concerning social life. Even in communities characterised by a certain cultural continuity, heated debates and arguments often arise, making it difficult to produce shared reflections and political solutions aimed at seeking what is good and just. Thus aside from the complexity of legitimate points of view found within the human family, there is also a factor emerging that seems to characterise the above-mentioned social situation, namely, a loss, or at least an eclipse, of the sense of what is human and an apparent reduction in the significance of the concept of human dignity.[12] Indeed, we seem to be losing the value and profound meaning of one of the fundamental concepts of the West: that of the human person. Thus, at a time when artificial intelligence programs are examining human beings and their actions, it is precisely the *ethos* concerning the understanding of the value and dignity of the human person that is most at risk in the implementation and development of these systems. Indeed, we must remember that no innovation is neutral. Technology is born for a purpose and, in its impact on human society, always represents a form of order in social relations and an arrangement of power, thus enabling certain people to perform specific actions while preventing others from performing different ones. In a more or less explicit way, this constitutive power

dimension of technology always includes the worldview of those who invented and developed it.

This likewise applies to artificial intelligence programs. In order for them to be instruments for building up the good and a better tomorrow, they must always be aimed at the good of every human being. They must have an ethical “inspiration”.

Moreover, an ethical decision is one that takes into account not only an action’s outcomes but also the values at stake and the duties that derive from those values. That is why I welcomed both the 2020 signing in Rome of the *Rome Call for AI Ethics*,[13] and its support for that type of ethical moderation of algorithms and artificial intelligence programs that I call “algor-ethics”.[14] In a pluralistic and global context, where we see different sensitivities and multiple hierarchies in the scales of values, it might seem difficult to find a single hierarchy of values. Yet, in ethical analysis, we can also make use of other types of tools: if we struggle to define a single set of global values, we can, however, find shared principles with which to address and resolve dilemmas or conflicts regarding how to live.

This is why the *Rome Call* was born: with the term “algor-ethics”, a series of principles are condensed into a global and pluralistic platform that is capable of finding support from cultures, religions, international organizations and major corporations, which are key players in this development.

The politics that is needed

We cannot, therefore, conceal the concrete risk, inherent in its fundamental design, that artificial intelligence might limit our worldview to realities expressible in numbers and enclosed in predetermined categories, thereby excluding the contribution of other forms of truth and imposing uniform anthropological, socio-economic and cultural models. The technological paradigm embodied in artificial intelligence runs the risk, then, of becoming a far more dangerous paradigm, which I have already identified as the “technocratic paradigm”.[15] We cannot allow a tool as powerful and indispensable as artificial intelligence to reinforce such a paradigm, but rather, we must make artificial intelligence a bulwark against its expansion.

This is precisely where political action is urgently needed. The Encyclical *Fratelli Tutti* reminds us that “for many people today, politics is a distasteful word, often due to the mistakes, corruption and inefficiency of some politicians. There are also attempts to discredit politics, to replace it with economics or to twist it to one ideology or another. Yet can our world function without politics? Can there be an effective process of growth towards universal fraternity and social peace without a sound political life?”.[16]

Our answer to these questions is: No! Politics is necessary! I want to reiterate in this moment that “in the face of many petty forms of politics focused on immediate interests [...] ‘true statecraft is manifest when, in difficult times, we uphold high principles and think of the long-term common good. Political powers do not find it easy to assume this duty in the work of nation-building’ (*Laudato Si’*, 178), much less in forging a common project for the human family, now and in the future”.[17]

Esteemed ladies and gentlemen!

My reflection on the effects of artificial intelligence on humanity leads us to consider the importance of “healthy politics” so that we can look to our future with hope and confidence. I have written previously that “global society is suffering from grave structural deficiencies that cannot be resolved by piecemeal solutions or quick fixes. Much needs to change, through fundamental reform and major renewal. Only a healthy politics, involving the most diverse sectors and skills, is capable of overseeing this process. An economy that is an integral part of a political, social, cultural and popular programme directed to the common good could pave the way for ‘different possibilities which do not involve stifling human creativity and its ideals of progress, but rather directing that energy along new channels’ (*Laudato Si’*, 191)”.[18]

This is precisely the situation with artificial intelligence. It is up to everyone to make good use of it but the onus is

on politics to create the conditions for such good use to be possible and fruitful.

Thank you.

_____

[1] *Message for the 57th World Day of Peace*, 1 January 2024, 1.

[2] Cf. *ibid*.

[3] Cf. *ibid.*, 2.

[4] This ambivalence was already noted by Pope Saint Paul VI in his *Address to the Personnel of the "Centro Automazione Analisi Linguistica" of the Aloysianum*, 19 June 1964.

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

[6] Cf. Encyclical Letter *Laudato Si'* (24 May 2015), 102-114.

[7] *Message for the 57th World Day of Peace*, 1 January 2024, 3.

[8] The insights of Marshall McLuhan and John M. Culkin are especially relevant to the consequences of the use of artificial intelligence.

[9] Cf. *Address to Participants in the Plenary Assembly of the Pontifical Academy for Life*, 28 February 2020.

[10] Cf. *Message for the 57th World Day of Peace*, 1 January 2024, 4.

[11] Cf. *ibid.*, 3, 7.

[12] Cf. Dicastery for the Doctrine of the Faith, Declaration *Dignitas Infinita* on Human Dignity (2 April 2024).

[13] Cf. *Address to Participants in the Plenary Assembly of the Pontifical Academy for Life*, 28 February 2020.

[14] Cf. *Address to Participants in the Congress on Child Dignity in the Digital World,* 14 November 2019; *Address to Participants in the Plenary Assembly of the Pontifical Academy for Life*, 28 February 2020.

[15] For a more extensive explanation, see the Encyclical Letter *Laudato Si'* on Care for Our Common Home (24 May 2015).

[16] Encyclical Letter, *Fratelli Tutti* on Fraternity and Social Friendship (3 October 2020), 176.

[17] *Ibid*, 178.

[18] *Ibid*, 179.

[01030-EN.01] [Original text: Italian]

Traduzione in lingua tedesca

Ein faszinierendes und unheimliches Instrument

Sehr geehrte Damen und Herren,

heute wende ich mich mit einigen Überlegungen zu den Auswirkungen der künstlichen Intelligenz auf die Zukunft der Menschheit an Sie, die Staats- und Regierungschefs des intergouvernementalen G7-Forums.

»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)«[1]. Wissenschaft und Technik sind also einzigartige Ergebnisse des kreativen Potenzials von uns Menschen[2].

Nun, genau aus der Nutzung dieses kreativen Potentials, das Gott uns gegeben hat, geht die künstliche Intelligenz hervor.

Letztere ist bekanntlich ein äußerst mächtiges Instrument, das in sehr vielen Bereichen des menschlichen Wirkens eingesetzt wird: von der Medizin bis zur Arbeitswelt, von der Kultur bis zur Kommunikation, von der Bildung bis zur Politik. Und man kann heute davon ausgehen, dass ihr Einsatz die Art und Weise, wie wir leben, unsere sozialen Beziehungen und in Zukunft sogar die Art und Weise, wie wir unsere Identität als Mensch begreifen, zunehmend beeinflussen wird[3].

Das Thema der künstlichen Intelligenz wird jedoch oft als ambivalent wahrgenommen: Einerseits begeistert es wegen der Möglichkeiten, die es bietet, andererseits flößt es Angst ein wegen der Konsequenzen, die es vorausahnen lässt. In diesem Zusammenhang kann man sagen, dass uns alle, wenn auch in unterschiedlichem Maße, zwei Gefühlslagen erfassen: Wir sind begeistert, wenn wir uns den Fortschritt vorstellen, der durch die künstliche Intelligenz erzielt werden kann, aber zugleich bekommen wir Angst, wenn wir die Gefahren bemerken, die mit ihrem Einsatz verbunden sind[4].

Wir können im Übrigen nicht daran zweifeln, dass das Aufkommen der künstlichen Intelligenz eine wahrhaft kognitiv-industrielle Revolution darstellt, die zur Schaffung eines neuen Gesellschaftssystems beitragen wird, das durch komplexe epochale Veränderungen gekennzeichnet ist. So könnte die künstliche Intelligenz beispielsweise eine Demokratisierung des Zugangs zu Wissen, den exponentiellen Fortschritt der wissenschaftlichen Forschung und die Möglichkeit, ermüdende Arbeiten an Maschinen abzugeben, ermöglichen; zugleich könnte sie aber auch eine größere Ungerechtigkeit zwischen fortgeschrittenen und sich entwickelnden Nationen, zwischen herrschenden und unterdrückten sozialen Schichten mit sich bringen und so die Möglichkeit einer „Kultur der Begegnung“ zugunsten einer „Kultur der Wegwerfens“ gefährden.

Das Ausmaß dieser komplexen Veränderungen ist offensichtlich mit der rasanten technologischen Entwicklung der künstlichen Intelligenz selbst verbunden.

Genau dieser rasante technologische Fortschritt macht die künstliche Intelligenz zu *einem faszinierenden und zugleich unheimlichen Instrument* und verlangt nach einer Reflexion, die der Situation gerecht wird.

In dieser Hinsicht könnte man vielleicht von der Feststellung ausgehen, dass die künstliche Intelligenz in erster Linie *ein Instrument* ist. Und es versteht sich von selbst, dass die Vorteile oder die Schäden, die sie mit sich bringen wird, von ihrem Einsatz abhängen.

Das ist sicher richtig, denn das war bei jedem Werkzeug der Fall, das der Mensch seit Anbeginn der Zeit hergestellt hat.

Diese unsere Fähigkeit, Werkzeuge in einer Menge und Komplexität herzustellen, die unter den Lebewesen einzigartig ist, lässt uns von einem *technisch-menschlichen Zustand* sprechen: Der Mensch hat schon immer eine Beziehung zur Umwelt gehabt, die durch die Werkzeuge vermittelt wurde, die er nach und nach hergestellt hat. Es ist unmöglich, die Geschichte des Menschen und der Zivilisation von der Geschichte dieser Werkzeuge

zu trennen. Manche haben in all dem eine Art Mangel, ein Defizit des Menschen sehen wollen, so als ob er aufgrund dieses Mangels gezwungen wäre, die Technik zu erfinden[5]. Ein aufmerksamer und objektiver Blick zeigt uns jedoch das Gegenteil. Wir leben im Hinblick auf unser biologisches Sein in einem Zustand des Darüberhinausgehens; wir sind Wesen, die nach außen hin exponiert, ja radikal offen sind für das Darüberhinaus. Darin hat unsere Offenheit für andere und für Gott ihren Ursprung; von dort geht das schöpferische Potential unserer Intelligenz in Form von Kultur und Schönheit hervor; und daraus entstehen schließlich unsere technischen Fähigkeiten. Die Technologie ist also eine Spur dieses unseres Darüberhinausgehens.

Der Gebrauch unserer Werkzeuge ist jedoch nicht immer eindeutig auf das Gute ausgerichtet. Auch wenn der Mensch in sich eine Berufung zum Darüberhinaus und zur Erkenntnis verspürt, die als Werkzeug des Guten im Dienst der Brüder und Schwestern und des gemeinsamen Hauses gelebt wird (vgl. *Gaudium et spes*, 16), so geschieht dies nicht immer. Im Gegenteil, die Menschheit hat gerade wegen ihrer radikalen Freiheit die Ziele ihres Daseins nicht selten pervertiert, indem sie sich selbst und dem Planeten zum Feind wurde[6]. Dasselbe Schicksal kann auch die technischen Mittel ereilen. Nur wenn ihre Berufung zum Dienst am Menschen gewährleistet ist, werden die technischen Mittel nicht nur die einzigartige Größe und Würde des Menschen offenbaren, sondern auch den Auftrag, den dieser erhalten hat, den Planeten und alle seine Bewohner „zu bearbeiten und zu hüten“ (vgl. *Gen* 2,15). Wenn wir über Technologie sprechen, sprechen wir darüber, was es bedeutet, Mensch zu sein, und damit über unseren einzigartigen Zustand zwischen Freiheit und Verantwortung; wir sprechen also über Ethik.

Als unsere Vorfahren nämlich Feuersteine schärften, um daraus Messer zu machen, benutzten sie diese, sowohl um Leder für Kleidung zuzuschneiden als auch um sich gegenseitig zu töten. Das Gleiche könnten wir über andere, viel fortschrittlichere Technologien sagen, wie zum Beispiel die Energie, die durch die Kernfusion erzeugt wird, wie es auf der Sonne geschieht, die sicher dafür genutzt werden könnte, um saubere, erneuerbare Energie zu erzeugen, aber auch um unseren Planeten in Schutt und Asche zu verwandeln.

Künstliche Intelligenz ist jedoch ein noch komplexeres Werkzeug. Ich würde fast sagen, dass es sich um ein Werkzeug *sui generis* handelt. Während also die Verwendung eines einfachen Werkzeugs (wie z. B. eines Messers) unter der Kontrolle des Menschen steht, der es benutzt, und nur von ihm seine ordnungsgemäße Verwendung abhängt, kann sich die künstliche Intelligenz dagegen autonom an die ihr zugewiesene Aufgabe anpassen und, wenn sie so konzipiert ist, unabhängig vom Menschen eine Auswahl treffen, um das gesetzte Ziel zu erreichen[7].

Es sollte immer bedacht werden, dass die Maschine in einigen Formen und mit diesen neuen Mitteln algorithmische Auswahlen treffen kann. Was die Maschine tut, ist eine technische Auswahl unter mehreren Möglichkeiten und beruht entweder auf genau definierten Kriterien oder auf statistischen Rückschlüssen. Der Mensch hingegen wählt nicht nur aus, sondern ist in seinem Herzen zu einer Entscheidung fähig. Die Entscheidung ist ein Element, das man eher als strategisch denn als eine Auswahl bezeichnen könnte, und sie erfordert eine praktische Bewertung. Zuweilen, wie so oft in der schwierigen Aufgabe des Regierens, sind wir gerufen, Entscheidungen zu treffen, die Konsequenzen für viele Menschen haben. Das menschliche Nachsinnen hat in diesem Zusammenhang immer von Weisheit gesprochen, die *Phronesis* der griechischen Philosophie und zumindest teilweise die Weisheit der Heiligen Schrift. Angesichts der Wunderwerke der Maschinen, die in der Lage zu sein scheinen, unabhängig auszuwählen, müssen wir uns darüber im Klaren sein, dass die Entscheidung immer dem Menschen überlassen bleiben muss, selbst in den dramatischen und dringlichen Situationen, bei denen sie manchmal in unserem Leben auf uns zukommt. Wir würden die Menschheit zu einer hoffnungslosen Zukunft verdammen, wenn wir den Menschen die Fähigkeit nehmen würden, über sich selbst und ihr Leben zu entscheiden, und sie dazu verdammen würden, von der Wahl von Maschinen abhängig zu sein. Wir müssen der menschlichen Kontrolle über den Auswahlprozess von Programmen der künstlichen Intelligenz einen bedeutenden Raum geben, diesen garantieren und schützen: Die menschliche Würde selbst steht dabei auf dem Spiel.

Gerade in dieser Frage möchte ich darauf bestehen, dass es in einem Drama wie einem bewaffneten Konflikt dringend erforderlich ist, die Entwicklung und den Gebrauch von Geräten wie den so genannten „tödlichen autonomen Waffen“ zu überdenken, um ihren Einsatz zu verbieten, beginnend mit einer proaktiven und

konkreten Verpflichtung zur Einführung einer immer größeren und bedeutenden menschlichen Kontrolle. Keine Maschine darf jemals die Wahl treffen können, einem Menschen das Leben zu nehmen.

Hinzu kommt, dass zumindest bei den fortgeschrittenen Formen der künstlichen Intelligenz die ordnungsgemäße Nutzung weder von den Nutzern noch von den Programmierern, die den ursprünglichen Zweck der Programme zum Zeitpunkt ihrer Entwicklung festgelegt haben, vollständig kontrolliert werden kann. Dies gilt umso mehr, als es sehr wahrscheinlich ist, dass in nicht allzu ferner Zukunft Programme künstlicher Intelligenz in der Lage sein werden, direkt miteinander zu kommunizieren, um ihre *Performance* zu verbessern. Und wenn in der Vergangenheit die Menschen, die einfache Werkzeuge modellierten, sahen, wie diese ihre Existenz prägten – das Messer ermöglichte ihnen das Überleben in der Kälte, aber auch die Entwicklung der Kriegskunst –, so werden sie jetzt, da sie ein komplexes Werkzeug modelliert haben, sehen, wie dieses ihre Existenz noch mehr prägt[8].

Der grundlegende Mechanismus der künstlichen Intelligenz

Ich möchte nun kurz auf die Komplexität der künstlichen Intelligenz eingehen. Künstliche Intelligenz ist im Wesentlichen ein Werkzeug zur Lösung eines Problems und funktioniert durch eine logische Verkettung von algebraischen Operationen, die mit Datenkategorien durchgeführt werden, die verglichen werden, um Korrelationen zu entdecken und ihren statistischen Wert zu verbessern, dank eines Selbstlernprozesses, der auf der Suche nach weiteren Daten und der Selbstmodifikation ihrer Berechnungsverfahren beruht.

Die künstliche Intelligenz ist also darauf ausgelegt, spezifische Probleme zu lösen, aber für diejenigen, die sie nutzen, ist die Versuchung oft unwiderstehlich, aus den von ihr vorgeschlagenen spezifischen Lösungen allgemeine, sogar anthropologische Schlüsse zu ziehen.

Ein gutes Beispiel ist der Einsatz von Programmen, die Richtern bei Entscheidungen über die Gewährung von Hausarrest für Verurteilte, die eine Strafe in einer Haftanstalt verbüßen, helfen sollen. In diesem Fall wird die künstliche Intelligenz verwendet, um die Wahrscheinlichkeit der Rückfälligkeit einer von einem Verurteilten begangenen Straftat anhand vordefinierter Kategorien (Art der Straftat, Verhalten im Gefängnis, psychologische Bewertung und andere) vorherzusagen, wobei die künstliche Intelligenz Zugang zu Datenkategorien hat, die das Privatleben des Verurteilten betreffen (ethnische Herkunft, Bildungsniveau, Kreditlinie und andere). Die Anwendung einer solchen Methodik – die mitunter das Risiko birgt, einer Maschine *de facto* das letzte Wort über das Schicksal einer Person zu überlassen – kann implizit eine Bezugnahme auf die vorgefassten Bewertungen mit sich bringen, die den von der künstlichen Intelligenz verwendeten Datenkategorien innewohnen.

Die Zugehörigkeit zu einer bestimmten ethnischen Gruppe oder, prosaischer ausgedrückt, eine geringfügige Ordnungswidrigkeit, die Jahre zuvor begangen wurde (z. B. das Nichtbezahlen eines Strafzettels), wird die Entscheidung, ob Hausarrest gewährt wird oder nicht, tatsächlich beeinflussen. Dagegen entwickelt sich der Mensch ständig weiter und ist in der Lage, mit seinen Handlungen zu überraschen, was die Maschine nicht berücksichtigen kann.

Es ist auch anzumerken, dass ähnliche Anwendungen wie die soeben erwähnte dadurch beschleunigt werden, dass Programme der künstlichen Intelligenz zunehmend mit der Fähigkeit ausgestattet werden, direkt mit Menschen zu interagieren (*Chatbots*), Gespräche mit ihnen zu führen und enge Beziehungen zu ihnen aufzubauen, die oft sehr angenehm und beruhigend scheinen, da diese Programme der künstlichen Intelligenz so konzipiert sind, dass sie lernen, in personalisierter Form auf die physischen und psychischen Bedürfnisse der Menschen zu reagieren.

Zu vergessen, dass die künstliche Intelligenz nicht ein menschliches Wesen ist und dass sie keine allgemeinen Prinzipien vorschlagen kann, ist oft ein schwerer Fehler, der entweder auf das tiefe Bedürfnis des Menschen zurückzuführen ist, eine dauerhafte Form der Gemeinschaft zu finden, oder auf eine unterbewusste Annahme des Menschen, nämlich die Annahme, dass Beobachtungen, die durch einen Rechenmechanismus gewonnen werden, unbestrittenen sicher und zweifellos allgemeingültig sind.

Diese Annahme ist jedoch sehr gewagt, wie eine Untersuchung der inhärenten Grenzen des Rechnens zeigt. Die künstliche Intelligenz verwendet algebraische Operationen, die in einer logischen Reihenfolge ausgeführt werden (z. B. wenn der Wert von X größer ist als der von Y, wird X mit Z multipliziert; andernfalls wird X durch Y geteilt). Diese Berechnungsmethode – der so genannte „Algorithmus“ – ist weder objektiv noch neutral[9]. Da sie auf Algebra basiert, kann sie nur numerisch formalisierte Wirklichkeiten untersuchen[10].

Es sollte auch nicht vergessen werden, dass die Algorithmen zur Lösung sehr komplexer Probleme so ausgeklügelt sind, dass es für die Programmierer selbst schwierig ist, genau zu verstehen, wie sie ihre Ergebnisse erzielen. Dieser Trend zur Verfeinerung dürfte sich mit der Einführung von Quantencomputern, die nicht mit binären Schaltkreisen (Halbleitern oder Mikrochips), sondern nach den recht komplexen Gesetzen der Quantenphysik arbeiten werden, noch erheblich beschleunigen. Andererseits ist die kontinuierliche Einführung von immer leistungsfähigeren Mikrochips bereits jetzt eine der Ursachen für die überwiegende Nutzung künstlicher Intelligenz durch die wenigen Nationen, die über sie verfügen.

Ob ausgeklügelt oder nicht, die Qualität der Antworten, die Programme der künstlichen Intelligenz geben, hängt letztlich von den Daten ab, die sie verwenden, und wie diese Daten strukturiert sind.

Abschließend möchte ich auf einen letzten Bereich hinweisen, in dem die Komplexität des Mechanismus der sogenannten generativen künstlichen Intelligenz (*Generative Artificial Intelligence*) deutlich zutage tritt. Niemand bezweifelt, dass es heute großartige Werkzeuge für den Wissenszugang gibt, die sogar das *self-learning* und das *self-tutoring* in einer Vielzahl von Bereichen ermöglichen. Viele von uns sind beeindruckt von den Anwendungen, die leicht im Netz verfügbar sind, um einen Text zu verfassen oder ein Bild zu einem beliebigen Thema zu erstellen. Besonders angetan davon sind Studenten, die davon unverhältnismäßig viel Gebrauch machen, wenn sie Arbeiten anfertigen müssen.

Diese Studenten, die sich oft viel besser als ihre Professoren mit dem Einsatz von künstlicher Intelligenz auskennen und daran gewöhnt sind, vergessen jedoch, dass die so genannte generative künstliche Intelligenz streng genommen nicht wirklich „kreativ“ ist. Sie durchsucht in Wahrheit *Big Data* nach Informationen und verpackt sie in dem von ihr gewünschten Stil. Sie entwickelt keine neuen Konzepte oder Analysen. Sie wiederholt die gefundenen und gibt ihnen eine ansprechende Form. Und je öfter sie einen Begriff oder eine Hypothese wiederholt vorfindet, desto mehr hält sie diese für legitim und gültig. Sie ist also nicht „generierend“, sondern „verstärkend“, in dem Sinne, dass sie bestehende Inhalte neu ordnet und zu ihrer Konsolidierung beiträgt, oft ohne zu prüfen, ob sie Fehler oder Vorurteile enthält.

Dies birgt nicht nur die Gefahr, *Fake News* zu legitimieren und den Vorteil einer vorherrschenden Kultur zu stärken, sondern untergräbt auch den Bildungsprozess *in nuce.* Die Bildung, die den Schülern die Möglichkeit einer authentischen Reflexion bieten sollte, läuft Gefahr, auf eine Wiederholung von Begriffen reduziert zu werden, die allein aufgrund ihrer ständigen Wiederkehr zunehmend als unanfechtbar bewertet werden[11].

Für gemeinsame Ethikrichtlinien die Würde des Menschen neu in den Mittelpunkt stellen

Dem bisher Gesagten soll nun eine allgemeinere Feststellung hinzugefügt werden. Die gegenwärtige Epoche der technologischen Innovation geht nämlich mit einer besonderen und noch nie dagewesenen sozialen Situation einher: In den großen Fragen des gesellschaftlichen Lebens wird es immer schwieriger, Einigungen zu erzielen. Selbst in Gemeinschaften, die sich durch eine gewisse kulturelle Kontinuität auszeichnen, kommt es häufig zu hitzigen Debatten und Konfrontationen, die gemeinsame Überlegungen und politische Lösungen auf der Suche nach dem Guten und Gerechten erschweren. Jenseits der Komplexität von legitimen Ansichten, die die menschliche Familie kennzeichnen, gibt es einen Faktor, der diesen verschiedenen Debatten gemein ist. Es ist ein Verlust oder zumindest eine Verdunkelung des Sinns für das Menschliche und eine scheinbare Bedeutungslosigkeit des Begriffes der Menschenwürde[12] zu verzeichnen. Es scheint, als würde der Wert und die tiefe Bedeutsamkeit einer der grundlegenden Kategorien des Westens verloren gehen: die Kategorie der menschlichen Person. In dieser Zeit, in der Programme der künstlichen Intelligenz den Menschen und sein Handeln in Frage stellen, stellt gerade der Mangel an einem *Ethos*, das mit der Wahrnehmung des Wertes und der Würde der menschlichen Person verbunden ist, den größten Schwachpunkt bei der Implementierung und

Entwicklung dieser Systeme dar. Wir dürfen nämlich nie vergessen, dass keine Innovation neutral ist. Die Technik wird zu einem bestimmten Zweck entwickelt und stellt in ihrer Auswirkung auf die menschliche Gesellschaft immer eine Form der Ordnung der sozialen Beziehungen und eine Machtmöglichkeit dar, die es den einen erlaubt, gewisse Handlungen auszuführen und die anderen daran hindert. Diese konstitutive Machtdimension der Technik trägt in sich immer, mehr oder weniger explizit, die Weltanschauung derjenigen eingeschrieben, die sie realisiert und entwickelt haben.

Dies gilt auch für Programme künstlicher Intelligenz. Damit diese zu Instrumenten für den Aufbau des Guten und einer besseren Zukunft werden können, müssen sie immer auf das Wohl jedes einzelnen Menschen ausgerichtet sein. Sie bedürfen einer ethischen Ausrichtung.

Die ethische Entscheidung ist in der Tat eine, die nicht nur die Ergebnisse einer Handlung berücksichtigt, sondern auch die Werte, die auf dem Spiel stehen, und die Pflichten, die sich aus diesen Werten ableiten. Aus diesem Grund habe ich die Unterzeichnung des „*Rome Call for AI Ethics*“[13] im Jahr 2020 in Rom und die darin enthaltene Unterstützung für diese Form der ethischen Moderation von Algorithmen und Programmen der künstlichen Intelligenz begrüßt, die ich als „Algor-Ethik“[14] bezeichnet habe. In einem pluralistischen und globalen Kontext, in dem sich auch unterschiedliche Wahrnehmungen und pluralistische Hierarchien in den Werteskalen zeigen, könnte es schwierig erscheinen, eine einzige Hierarchie der Werte zu finden. Aber in der ethischen Analyse können wir auch auf andere Arten von Instrumenten zurückgreifen: Wenn es uns schwerfällt, eine einzige Ordnung globaler Werte zu definieren, so können wir doch gemeinsame Prinzipien finden, mit denen wir etwaige Dilemmata oder Konflikte im Leben angehen und lösen können.

Aus diesem Grund wurde der *Rome Call* ins Leben gerufen: In dem Begriff „Algor-Ethik“ wird eine Reihe von Prinzipien zu einer globalen und pluralistischen Plattform verdichtet, die in der Lage ist, die Zustimmung von Kulturen, Religionen, internationalen Organisationen und großen Unternehmen als Hauptakteure dieser Entwicklung zu finden.

Die Politik, die wir dazu brauchen

Es besteht also die konkrete Gefahr, dass die künstliche Intelligenz gemäß ihrem grundlegenden Mechanismus die Sicht der Welt auf in Zahlen ausgedrückte und in vorgefertigte Kategorien gefasste Wirklichkeiten beschränkt, den Beitrag anderer Ausdrucksformen der Wahrheit verdrängt und einheitliche anthropologische, sozioökonomische und kulturelle Modelle aufzwingt. Das technologische Paradigma, das durch die künstliche Intelligenz verkörpert wird, läuft also Gefahr, einem weitaus gefährlicheren Paradigma Platz zu machen, das ich bereits als das „technokratische Paradigma“[15] beschrieben habe. Wir dürfen nicht zulassen, dass ein so mächtiges und unentbehrliches Werkzeug wie die künstliche Intelligenz ein solches Paradigma verstärkt; vielmehr müssen wir gerade die künstliche Intelligenz zu einem Schutzwall gegen seine Ausbreitung machen.

Und genau hier besteht dringender politischer Handlungsbedarf, wie die Enzyklika *Fratelli tutti* in Erinnerung ruft. Gewiss, »für viele ist die heutige Politik ein Schimpfwort, und es ist nicht zu übersehen, dass hinter dieser Tatsache oft Fehler, Korruption und Ineffizienz mancher Politiker stehen. Hierzu kommen noch Strategien, die darauf abzielen, die Politik zu schwächen, sie durch die Wirtschaft zu ersetzen oder sie mit einer Ideologie zu beherrschen. Und dennoch, kann die Welt ohne Politik funktionieren? Kann sie ohne eine gute Politik einen effektiven Weg zur allgemeinen Geschwisterlichkeit und zum gesellschaftlichen Frieden finden?«[16].

Unsere Antwort auf diese letzten Fragen lautet: Nein! Die Politik ist notwendig! Ich möchte bei dieser Gelegenheit noch einmal betonen: »Angesichts vieler Formen armseliger Politik, die auf das unmittelbare Interesse ausgerichtet sind, zeigt sich „die politische Größe [...], wenn man in schwierigen Momenten nach bedeutenden Grundsätzen handelt und dabei an das langfristige Gemeinwohl denkt. Diese Pflicht in einem Projekt der Nation auf sich zu nehmen, kostet die politische Macht einen hohen Preis“, dies umso mehr in einem gemeinsamen Projekt für die gegenwärtige und zukünftige Menschheit«[17].

Sehr geehrte Damen und Herren,

meine Überlegungen zu den Auswirkungen der künstlichen Intelligenz auf die Zukunft der Menschheit führen uns zu der Feststellung, wie wichtig eine „gesunde Politik“ ist, um mit Hoffnung und Zuversicht in die Zukunft zu blicken. Wie ich bereits an anderer Stelle gesagt habe: »Die weltweite Gesellschaft weist schwerwiegende strukturelle Mängel auf, die nicht durch Zusammenflicken oder bloße schnelle Gelegenheitslösungen behoben werden. Es gibt Dinge, die durch neue Grundausrichtungen und bedeutende Verwandlungen verändert werden müssen. Nur eine gesunde Politik könnte hier die Führungsrolle übernehmen und dabei die verschiedensten Sektoren und die unterschiedlichsten Wissensbereiche einbeziehen. So kann eine Wirtschaft, die sich in ein politisches, soziales, kulturelles und vom Volk her kommendes Projekt für das Gemeinwohl einfügt, „den Weg für andere Möglichkeiten [eröffnen], die nicht etwa bedeuten, die Kreativität des Menschen und seinen Sinn für Fortschritt zu bremsen, sondern diese Energie auf neue Anliegen hin auszurichten“ (*Laudato si’*, 191)«[18].

Genau das ist bei der künstlichen Intelligenz der Fall. Es liegt an allen, sie sinnvoll zu nutzen, und es kommt Politik zu, die Bedingungen dafür zu schaffen, dass eine solche positive Nutzung möglich und fruchtbar ist.

Danke.

---

[1] *Botschaft zum 57. Weltfriedenstag am 1. Januar 2024*, 1.

[2] Vgl. *ebd*.

[3] Vgl. *ebd.*, 2.

[4] Diese Ambivalenz wurde bereits von Papst Paul VI. in seiner *Ansprache an die Mitarbeiter des „Zentrums für die Automatisierung linguistischer Analyse“ des Aloysianum* am 19. Juni 1964 wahrgenommen.

[5] Vgl. A. Gehlen, *Der Mensch: Seine Natur und seine Stellung in der Welt*, Wiesbaden 142004, 16f.

[6] Enzyklika *Laudato si’* (24. Mai 2015), 102-114.

[7] *Botschaft zum 57. Weltfriedenstag am 1. Januar 2024*, 3.

[8] Die Erkenntnisse von Marshall McLuhan und John M. Culkin sind besonders relevant für die Folgen des Einsatzes künstlicher Intelligenz.

[9] Vgl. *Ansprache an die Teilnehmer der Vollversammlung der Päpstlichen Akademie für das Leben,* 28. Februar 2020.

[10] *Botschaft zum 57. Weltfriedenstag am 1. Januar 2024*, 4.

[11] Vgl. *ebd.*, 3 und 7.

[12] Vgl. Dikasterium für die Glaubenslehre, Erklärung *Dignitas infinita* über die menschliche Würde (2. April 2024).

[13] Vgl. *Ansprache an die Teilnehmer der Vollversammlung der Päpstlichen Akademie für das Leben*

(28. Februar 2020).

[14] Vgl. *Ansprache an die Teilnehmer des Kongresses „Promoting Digital Child Dignity – From Concept to*

Action“ (14. November 2019); vgl. *Ansprache an die Teilnehmer der Vollversammlung der Päpstlichen Akademie für das Leben* (28. Februar 2020).

[15] Für eine ausführlichere Darstellung verweise ich auf meine Enzyklika *Laudato si'* über die Sorge für das gemeinsame Haus (24. Mai 2015).

[16] Enzyklika *Fratelli tutti* über die Geschwisterlichkeit und soziale Freundschaft (3. Oktober 2020), 176.

[17] *Ebd.*, 178.

[18] *Ebd.*, 179.

[01030-DE.01] [Originalsprache: Italienisch]

Traduzione in lingua spagnola

Un instrumento fascinante y tremendo

Estimadas señoras, distinguidos señores:

Me dirijo hoy a ustedes, líderes del Foro Intergubernamental del G7, con una reflexión sobre los efectos de la inteligencia artificial en el futuro de la humanidad.

«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)»[1]. La ciencia y la tecnología son, por lo tanto, producto extraordinario del potencial creativo que poseemos los seres humanos[2].

Ahora bien, la inteligencia artificial se origina precisamente a partir del uso de este potencial creativo que Dios nos ha dado.

Dicha inteligencia artificial, como sabemos, es un instrumento extremadamente poderoso, que se emplea en numerosas áreas de la actividad humana: de la medicina al mundo laboral, de la cultura al ámbito de la comunicación, de la educación a la política. Y es lícito suponer, entonces, que su uso influirá cada vez más en nuestro modo de vivir, en nuestras relaciones sociales y en el futuro, incluso en la manera en que concebimos nuestra identidad como seres humanos[3].

El tema de la inteligencia artificial, sin embargo, a menudo es percibido de modo ambivalente: por una parte, entusiasma por las posibilidades que ofrece; por otra, provoca temor ante las consecuencias que podrían llegar a producirse. A este respecto podríamos decir que todos nosotros, aunque en diferente medida, estamos atravesados por dos emociones: somos entusiastas cuando imaginamos los progresos que se pueden derivar de la inteligencia artificial, pero, al mismo tiempo, nos da miedo cuando constatamos los peligros inherentes a su uso[4].

No podemos dudar, ciertamente, de que la llegada de la inteligencia artificial representa una auténtica revolución cognitiva-industrial, que contribuirá a la creación de un nuevo sistema social caracterizado por complejas transformaciones de época. Por ejemplo, la inteligencia artificial podría permitir una democratización del acceso al saber, el progreso exponencial de la investigación científica, la posibilidad de delegar a las máquinas los trabajos desgastantes; pero, al mismo tiempo, podría traer consigo una mayor inequidad entre naciones avanzadas y naciones en vías de desarrollo, entre clases sociales dominantes y clases sociales oprimidas, poniendo así en peligro la posibilidad de una “cultura del encuentro” y favoreciendo una “cultura del descarte”.

La magnitud de estas complejas transformaciones está vinculada obviamente al rápido desarrollo tecnológico de la misma inteligencia artificial.

Es precisamente este poderoso avance tecnológico el que hace de la inteligencia artificial *un instrumento fascinante* y *tremendo* al mismo tiempo, y exige una reflexión a la altura de la situación.

En esa dirección tal vez se podría partir de la constatación de que la inteligencia artificial es sobre todo *un instrumento*. Y resulta espontáneo afirmar que los beneficios o los daños que esta conlleve dependerán de su uso.

Esto es cierto, porque ha sido así con cada herramienta construida por el ser humano desde el principio de los tiempos.

Nuestra capacidad de construir herramientas, en una cantidad y complejidad que no tiene igual entre los seres vivos, nos habla de una *condición tecno-humana*. El ser humano siempre ha mantenido una relación con el ambiente mediada por los instrumentos que iba produciendo. No es posible separar la historia del hombre y de la civilización de la historia de esos instrumentos. Algunos han querido leer en todo eso una especie de privación, un déficit del ser humano, como si, a causa de esa carencia, estuviera obligado a dar vida a la tecnología[5]. Una mirada atenta y objetiva en realidad nos muestra lo contrario. Vivimos una condición de ulterioridad respecto a nuestro ser biológico; somos seres inclinados hacia el fuera-de-nosotros, es más, radicalmente abiertos al más allá. De aquí se origina nuestra apertura a los otros y a Dios; de aquí nace el potencial creativo de nuestra inteligencia en términos de cultura y de belleza; de aquí, por último, se origina nuestra capacidad técnica. La tecnología es así una huella de nuestra ulterioridad.

Sin embargo, el uso de nuestras herramientas no siempre está dirigido unívocamente al bien. Aun cuando el ser humano siente dentro de sí una vocación al más allá y al conocimiento vivido como instrumento de bien al servicio de los hermanos y hermanas, y de la *casa común* (cf. *Gaudium et spes*, 16), esto no siempre sucede. Es más, no pocas veces, precisamente gracias a su libertad radical, la humanidad ha pervertido los fines de su propio ser, transformándose en enemiga de sí misma y del planeta[6]. La misma suerte pueden correr los instrumentos tecnológicos. Solamente si se garantiza su vocación al servicio de lo humano, los instrumentos tecnológicos revelarán no sólo la grandeza y la dignidad única del ser humano, sino también el mandato que este último ha recibido de “cultivar y cuidar” el planeta y todos sus habitantes (cf. *Gn* 2,15). Hablar de tecnología es hablar de lo que significa ser humanos y, por tanto, de nuestra condición única entre libertad y responsabilidad, es decir, significa hablar de ética.

De hecho, cuando nuestros antepasados afilaron piedras de sílex para hacer cuchillos, los usaron tanto para cortar pieles para vestirse como para eliminarse entre sí. Lo mismo podría decirse de otras tecnologías mucho más avanzadas, como la energía producida por la fusión de los átomos, como ocurre en el Sol, que podría utilizarse para producir energía limpia y renovable, pero también para reducir nuestro planeta a cenizas.

Pero la inteligencia artificial es una herramienta aún más compleja. Yo diría que es una herramienta *sui generis*. Así, mientras que el uso de una herramienta simple —como un cuchillo— está bajo el control del ser humano que lo utiliza y su buen uso depende sólo de él, la inteligencia artificial, en cambio, puede adaptarse de forma autónoma a la tarea que se le asigne y, si se diseña de esa manera, podría tomar decisiones independientemente del ser humano para alcanzar el objetivo fijado [7].

Conviene recordar siempre que la máquina puede, en algunas formas y con estos nuevos medios, elegir por medio de algoritmos. Lo que hace la máquina es una elección técnica entre varias posibilidades y se basa en criterios bien definidos o en inferencias estadísticas. El ser humano, en cambio, no sólo elige, sino que en su corazón es capaz de decidir. La decisión es un elemento que podríamos definir el más estratégico de una elección y requiere una evaluación práctica. A veces, frecuentemente en la difícil tarea de gobernar, también estamos llamados a decidir con consecuencias para muchas personas. Desde siempre la reflexión humana habla a este propósito de sabiduría, la *phronesis* de la filosofía griega y, al menos en parte, la sabiduría de la Sagrada Escritura. Frente a los prodigios de las máquinas, que parecen saber elegir de manera independiente,

debemos tener bien claro que al ser humano le corresponde siempre la decisión, incluso con los tonos dramáticos y urgentes con que a veces ésta se presenta en nuestra vida. Condenaríamos a la humanidad a un futuro sin esperanza si quitáramos a las personas la capacidad de decidir por sí mismas y por sus vidas, condenándolas a depender de las elecciones de las máquinas. Necesitamos garantizar y proteger un espacio de control significativo del ser humano sobre el proceso de elección utilizado por los programas de inteligencia artificial. Está en juego la misma dignidad humana.

Precisamente sobre este tema, permítanme insistir en que, en un drama como el de los conflictos armados, es urgente replantearse el desarrollo y la utilización de dispositivos como las llamadas “armas autónomas letales” para prohibir su uso, empezando desde ya por un compromiso efectivo y concreto para introducir un control humano cada vez mayor y significativo. Ninguna máquina debería elegir jamás poner fin a la vida de un ser humano.

Hay que añadir, además, que el buen uso, al menos de las formas avanzadas de inteligencia artificial, no estará plenamente bajo el control ni de los usuarios ni de los programadores que definieron sus objetivos iniciales en el momento de elaborarlos. Y esto es tanto más cierto cuanto que es muy probable que, en un futuro no lejano, los programas de inteligencias artificiales puedan comunicarse directamente entre sí, para mejorar su rendimiento. Y, si en el pasado, los seres humanos que utilizaron herramientas simples vieron su existencia modelada por estos últimos —el cuchillo les permitió sobrevivir al frío pero también desarrollar el arte de la guerra—, ahora que los seres humanos han modelado un instrumento complejo, verán que este modelará aún más su existencia[8].

El mecanismo básico de la inteligencia artificial

Permítanme ahora detenerme brevemente sobre la complejidad de la inteligencia artificial. Básicamente, la inteligencia artificial es una herramienta diseñada para resolver un problema y funciona mediante un encadenamiento lógico de operaciones algebraicas, realizado en base a categorías de datos, que se comparan para descubrir correlaciones y mejorar su valor estadístico mediante un proceso de autoaprendizaje basado en la búsqueda de datos adicionales y la automodificación de sus procedimientos de cálculo.

La inteligencia artificial está diseñada de este modo para resolver problemas específicos, pero para quienes la utilizan la tentación de obtener, a partir de las soluciones puntuales que propone, deducciones generales, incluso de orden antropológico, es a menudo irresistible.

Un buen ejemplo es el uso de programas diseñados para ayudar a los magistrados en las decisiones relativas a la concesión de prisión domiciliaria a presos que están cumpliendo una condena en una institución penitenciaria. En este caso, se pide a la inteligencia artificial que prevea la probabilidad de reincidencia del delito cometido por un condenado a partir de categorías prefijadas (tipo de delito, comportamiento en prisión, evaluación psicológica y otros) lo que permite a la inteligencia artificial tener acceso a categorías de datos relacionados con la vida privada de la persona detenida (origen étnico, nivel educativo, línea de crédito, etc.). El uso de tal metodología —que a veces corre el riesgo de delegar *de facto* en una máquina la última palabra sobre el destino de una persona— puede llevar implícitamente la referencia a los prejuicios inherentes a las categorías de datos utilizados por la inteligencia artificial.

El ser clasificado en un cierto grupo étnico o, más prosaicamente, el haber cometido hace años una pequeña infracción —el no haber pagado, por ejemplo, una multa por aparcar en zona prohibida—, influirá, de hecho, en la decisión acerca de la concesión de la prisión domiciliaria. Por el contrario, el ser humano está siempre en evolución y es capaz de sorprender con sus acciones, algo que la máquina no puede tener en cuenta.

Hay que evidenciar también que aplicaciones análogas a ésta de la que estamos hablando se multiplicarán gracias al hecho de que los programas de inteligencia artificial estarán cada vez más dotados de la capacidad de interactuar directamente con los seres humanos (*chatbots*), sosteniendo conversaciones y estableciendo relaciones de cercanía con ellos, con frecuencia muy agradables y tranquilizadoras, en cuanto tales programas de inteligencia artificial están diseñados para aprender a responder, de forma personalizada, a las necesidades

físicas y psicológicas de los seres humanos.

Olvidar que la inteligencia artificial no es otro ser humano y que no puede proponer principios generales, es a veces un gran error que parte de la profunda necesidad de los seres humanos de encontrar una forma estable de compañía, o bien de un presupuesto subconsciente, es decir, de la creencia de que las observaciones obtenidas mediante un mecanismo de cálculo estén dotadas de las cualidades de certeza indiscutible y de universalidad indudable.

Esta suposición es, sin embargo, descabellada, como demuestra el examen de los límites intrínsecos del cálculo mismo. La inteligencia artificial usa operaciones algebraicas que se realizan según una secuencia lógica (por ejemplo, si el valor de X es superior al de Y, multiplica X por Y; si no divide X por Y). Este método de cálculo —denominado algoritmo— no está dotado ni de objetividad ni de neutralidad[9]. Al estar basado en el álgebra puede examinar sólo realidades formalizadas en términos numéricos[10].

No hay que olvidar, además, que los algoritmos diseñados para resolver problemas muy complejos son sofisticados de tal manera que hacen muy difícil a los propios programadores la comprensión exacta de cómo estos sean capaces de alcanzar sus resultados. Esta tendencia a la sofisticación corre el riesgo de acelerarse notablemente con la introducción de los ordenadores cuánticos que no operan con circuitos binarios (semiconductores o microchips), sino según las leyes, bastante articuladas, de la física cuántica. Por otra parte, la continua introducción de microchips cada vez más eficaces es la causa del predominio del uso de la inteligencia artificial por parte de las pocas naciones que disponen de ella.

La calidad de las respuestas que los programas de inteligencia artificial pueden dar, sean más o menos sofisticadas, depende en última instancia de los datos que manejan y de cómo estos los estructuran.

Finalmente, me gustaría señalar un último ámbito en el que emerge claramente la complejidad del mecanismo de la llamada inteligencia artificial generativa (*Generative Artificial Inteligence*). Nadie duda de que hoy en día están a disposición magníficos instrumentos de acceso al conocimiento que permiten incluso el autoaprendizaje (*self-learning*) y la autotutoría (*self-tutoring*) en una gran cantidad de campos. Muchos de nosotros nos hemos quedado sorprendidos por las aplicaciones fácilmente accesibles en línea para componer un texto o producir una imagen sobre cualquier tema o materia. Esto atrae de forma especial a los estudiantes que, cuando deben preparar los trabajos, hacen un uso desmedido.

Estos alumnos, que a menudo están mucho más preparados y acostumbrados al uso de la inteligencia artificial que sus profesores, olvidan, sin embargo, que la denominada inteligencia artificial generativa, en sentido estricto, no es propiamente “generativa”. En realidad, lo que esta hace es buscar información en los macrodatos (*big data*) y confeccionarla en el estilo que se le ha pedido. No desarrolla conceptos o análisis nuevos. Repite lo que encuentra, dándole una forma atractiva. Y cuanto más repetida encuentra una noción o una hipótesis, más la considera legítima y válida. Más que “generativa”, se la podría llamar “reforzadora”, en el sentido de que reordena los contenidos existentes, contribuyendo a consolidarlos, muchas veces sin controlar si tienen errores o prejuicios.

De este modo, no sólo se corre el riesgo de legitimar la difusión de noticias falsas y robustecer la ventaja de una cultura dominante, sino de minar también el proceso educativo en ciernes (*in nuce*). La educación, que debería dar a los estudiantes la posibilidad de una reflexión auténtica, corre el riesgo de reducirse a una repetición de nociones, que se considerarán cada vez más incontestables, simplemente a causa de ser continuamente presentadas[11].

Poner de nuevo al centro la dignidad de la persona en vista de una propuesta ética compartida

A lo que ya hemos dicho se añade una observación más general. La época de innovación tecnológica que estamos atravesando, en efecto, se acompaña de una particular e inédita coyuntura social, en la que cada vez es más difícil encontrar puntos de encuentro sobre los grandes temas de la vida social. Incluso en comunidades caracterizadas por una cierta continuidad cultural, se crean con frecuencia encendidos debates y choques que

hacen difícil llegar a acuerdos y soluciones políticas compartidas, orientadas a la búsqueda de lo que es bueno y justo. Además de la complejidad de las legítimas visiones que caracterizan a la familia humana, emerge un factor que parece acomunar estas distintas instancias. Se registra una pérdida o al menos un oscurecimiento del sentido de lo humano y una aparente insignificancia del concepto de dignidad humana[12]. Pareciera que se está perdiendo el valor y el profundo significado de una de las categorías fundamentales de Occidente: la categoría de persona humana. Y es así que en esta época en la que los programas de inteligencia artificial cuestionan al ser humano y su actuar, precisamente la debilidad del *ethos* vinculada a la percepción del valor y de la dignidad de la persona humana corre el riesgo de ser el mayor daño (*vulnus*) en la implementación y el desarrollo de estos sistemas. No debemos olvidar que ninguna innovación es neutral. La tecnología nace con un propósito y, en su impacto en la sociedad humana, representa siempre una forma de orden en las relaciones sociales y una disposición de poder, que habilita a alguien a realizar determinadas acciones impidiéndoselo a otros. Esta dimensión de poder que es constitutiva de la tecnología incluye siempre, de una manera más o menos explícita, la visión del mundo de quien la ha realizado o desarrollado.

Esto vale también para los programas de inteligencia artificial. Con el fin de que estos instrumentos sean para la construcción del bien y de un futuro mejor, deben estar siempre ordenados al bien de todo ser humano. Deben contener una inspiración ética.

La decisión ética, de hecho, es aquella que tiene en cuenta no sólo los resultados de una acción, sino también los valores en juego y los deberes que se derivan de esos valores. Por esto he acogido con satisfacción la firma en Roma, en 2020, de la *Rome Call for AI Ethics*[13] y su apoyo a esa forma de moderación ética de los algoritmos y de los programas de inteligencia artificial que he llamado “algorética”[14]. En un contexto plural y global, en el que también se muestran las distintas sensibilidades y plurales jerarquías en las escalas de valores, parecería difícil encontrar una única jerarquía de valores. Pero en el análisis ético podemos recurrir además a otros tipos de instrumentos. Si nos cuesta definir un solo conjunto de valores globales, podemos encontrar principios compartidos con los cuales afrontar y disminuir eventuales dilemas y conflictos de la vida.

Por esta razón ha nacido la *Rome Call*. En el término “algorética” se condensa una serie de principios que se revelan como una plataforma global y plural capaz de encontrar el apoyo de las culturas, las religiones, las organizaciones internacionales y las grandes empresas protagonistas de este desarrollo.

La política que se necesita

No podemos, por tanto, ocultar el riesgo concreto, porque es inherente a su mecanismo fundamental, de que la inteligencia artificial limite la visión del mundo a realidades que pueden expresarse en números y encerradas en categorías preestablecidas, eliminando la aportación de otras formas de verdad e imponiendo modelos antropológicos, socioeconómicos y culturales uniformes. El paradigma tecnológico encarnado por la inteligencia artificial corre el riesgo de dar paso a un paradigma mucho más peligroso, que ya he identificado con el nombre de “paradigma tecnocrático”[15]. No podemos permitir que una herramienta tan poderosa e indispensable como la inteligencia artificial refuerce tal paradigma, sino que más bien debemos hacer de la inteligencia artificial un baluarte precisamente contra su expansión.

Y es precisamente aquí donde urge la acción política, como recuerda la encíclica *Fratelli tutti*. Ciertamente «para muchos la política hoy es una mala palabra, y no se puede ignorar que detrás de este hecho están a menudo los errores, la corrupción, la ineficiencia de algunos políticos. A esto se añaden las estrategias que buscan debilitarla, reemplazarla por la economía o dominarla con alguna ideología. Pero, ¿puede funcionar el mundo sin política? ¿Puede haber un camino eficaz hacia la fraternidad universal y la paz social sin una buena política?»[16].

Nuestra respuesta a estas últimas preguntas es: ¡no! ¡La política sirve! Quiero reiterar en esta ocasión que «ante tantas formas mezquinas e inmediatistas de política [...], la grandeza política se muestra cuando, en momentos difíciles, se obra por grandes principios y pensando en el bien común a largo plazo. Al poder político le cuesta mucho asumir este deber en un proyecto de nación y más aún en un proyecto común para la humanidad presente y futura»[17].

Estimadas señoras, distinguidos señores:

Mi reflexión sobre los efectos de la inteligencia artificial en el futuro de la humanidad nos lleva así a la consideración de la importancia de la “sana política” para mirar con esperanza y confianza nuestro futuro. Como he dicho en otra ocasión, «la sociedad mundial tiene serias fallas estructurales que no se resuelven con parches o soluciones rápidas meramente ocasionales. Hay cosas que deben ser cambiadas con replanteos de fondo y transformaciones importantes. Sólo una sana política podría liderarlo, convocando a los más diversos sectores y a los saberes más variados. De esa manera, una economía integrada en un proyecto político, social, cultural y popular que busque el bien común puede “abrir camino a oportunidades diferentes, que no implican detener la creatividad humana y su sueño de progreso, sino orientar esa energía con cauces nuevos” (*Laudato si’*, 191)»[18].

Este es precisamente el caso de la inteligencia artificial. Corresponde a cada uno hacer un buen uso de ella, y corresponde a la política crear las condiciones para que ese buen uso sea posible y fructífero.

Gracias.

---

[1] *Mensaje para la 57 Jornada Mundial de la Paz* (1 enero 2024), 1.

[2] Cf. *ibíd*.

[3] Cf. *ibíd*., 2.

[4] Esta ambivalencia ya había sido advertida por el Papa san Pablo VI en su *Discurso al personal del “Centro de Automación de Análisis Lingüísticos” del Aloisiano de Gallarate* (19 junio 1964).

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

[6] Carta enc. *Laudato si’* sobre el cuidado de la casa común (24 mayo 2015), 102-114.

[7] Cf. *Mensaje para la 57 Jornada Mundial de la Paz* (1 enero 2024), 3.

[8] Las ideas de Marshall McLuhan y John M. Culkin son particularmente relevantes para comprender las consecuencias del uso de la inteligencia artificial.

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

[10] Cf. *Mensaje para la 57 Jornada Mundial de la Paz* (1 enero 2024), 4.

[11] Cf. *ibíd*., 3 y 7.

[12] Cf. Dicasterio para la Doctrina de la Fe, Declaración *Dignitas infinita* sobre la dignidad humana (2 abril 2024).

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

[14] Cf. *Discurso a los participantes en el Congreso “Promoting Digital Child Dignity – From Concept to Action”* (14 noviembre 2019); *Discurso a los participantes en la Plenaria de la Pontificia Academia para la Vida* (28 febrero 2020).

[15] Para una exposición más amplia, remito a mi Carta encíclica *Laudato si'* sobre el cuidado de la casa común (24 mayo 2015).

[16] Carta enc. *Fratelli tutti* sobre la fraternidad y la amistad social (3 octubre 2020), 176.

[17] *Ibíd.*, 178.

[18] *Ibíd.*, 179.

[01030-ES.01] [Texto original: Italiano]

Traduzione in lingua portoghese

Um instrumento fascinante e tremendo

Prezadas Senhoras, Ilustres Senhores!

Dirijo-me hoje a vós, Líderes do Fórum Intergovernamental do G7, com uma reflexão sobre os efeitos da inteligência artificial no futuro da humanidade.

«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)»[1]. A ciência e a tecnologia são, por conseguinte, produtos extraordinários do nosso potencial criativo, como seres humanos[2].

Pois bem, a inteligência artificial emerge precisamente do uso deste potencial criativo que Deus nos deu.

Como é sabido, trata-se de um instrumento extremamente poderoso, utilizado em muitos domínios da atividade humana: da medicina ao mundo do trabalho, da cultura à comunicação, da educação à política. E é já legítimo supor que o seu uso influenciará cada vez mais a nossa forma de viver, as nossas relações sociais e, no futuro, até mesmo a maneira como concebemos a nossa identidade enquanto seres humanos[3].

No entanto, o tema da inteligência artificial é frequentemente percebido como ambivalente: por um lado, entusiasma pelas possibilidades que oferece; por outro, gera temor pelas consequências que deixa antever. A este respeito, pode dizer-se que todos nós somos, embora em graus diferentes, atravessados por duas emoções: ficamos entusiasmados quando imaginamos os progressos que podem advir da inteligência artificial, mas ao mesmo tempo amedrontados quando constatamos os perigos inerentes ao seu uso[4].

Além disso, não podemos duvidar que o advento da inteligência artificial represente uma verdadeira revolução cognitivo-industrial que contribuirá para a criação de um novo sistema social caracterizado por complexas transformações epocais. Por exemplo, a inteligência artificial poderia permitir uma democratização do acesso ao conhecimento, o progresso exponencial da investigação científica, a possibilidade de delegar às máquinas os trabalhos exaustivos; mas ao mesmo tempo, ela poderia trazer consigo uma maior injustiça entre nações desenvolvidas e nações em vias de desenvolvimento, entre classes sociais dominantes e classes sociais oprimidas, colocando em perigo a possibilidade de uma "cultura do encontro" em favor de uma "cultura do descarte".

O alcance dessas complexas transformações está obviamente ligado ao rápido desenvolvimento tecnológico da própria inteligência artificial.

Este vigoroso avanço tecnológico torna a inteligência artificial, simultaneamente, um *instrumento fascinante e tremendo*, e exige uma reflexão à altura da situação.

Nesse sentido, talvez se possa partir da constatação de que a inteligência artificial é, antes de tudo, *um instrumento*. E é natural afirmar que os benefícios ou danos que trará dependerão do modo como é utilizado.

Isso é certamente verdade, pois foi assim para cada ferramenta construída pelo ser humano desde o início dos tempos.

Essa nossa capacidade de construir utensílios numa quantidade e complexidade sem paralelo entre os seres vivos, faz-nos falar de uma *condição tecno-humana*: o ser humano sempre manteve uma relação com o ambiente mediada pelas ferramentas que ia produzindo. Não é possível separar a história do homem e da civilização da história desses instrumentos. Houve quem quisesse ler em tudo isto uma espécie de insuficiência, um *deficit* do ser humano, como se, por causa dessa carência, fosse obrigado a dar vida à tecnologia[5]. Na verdade, um olhar atento e objetivo mostra-nos o contrário. Vivemos uma condição de ulterioridade em relação ao nosso ser biológico; somos seres inclinados para fora-de-nós-mesmos, ou melhor, somos radicalmente abertos ao além. Aqui tem início a nossa abertura aos outros e a Deus; daqui nasce o potencial criativo da nossa inteligência em termos de cultura e beleza; e por fim, daqui se origina a nossa capacidade técnica. A tecnologia é, assim, uma marca desta nossa ulterioridade.

No entanto, o uso das nossas ferramentas nem sempre está orientado exclusivamente para o bem. Mesmo que o ser humano sinta interiormente uma vocação para além de si mesmo e para o conhecimento, vivido como instrumento de bem ao serviço dos irmãos, das irmãs e da *casa comum* (cf. *Gaudium et spes*, 16), isso nem sempre acontece. Ao contrário, não poucas vezes, precisamente graças à sua liberdade radical, a humanidade perverteu os fins do seu ser, transformando-se em inimiga de si mesma e do planeta[6]. O mesmo pode acontecer com os instrumentos tecnológicos. Somente se for garantida a sua vocação ao serviço do homem, os instrumentos tecnológicos revelarão não apenas a grandeza e a dignidade única do ser humano, mas também o mandato que este recebeu de “cultivar e guardar” (cfr. *Gn* 2,15) o planeta e todos os seus habitantes. Falar de tecnologia é falar sobre o que significa ser humano e, portanto, sobre aquela nossa condição única entre liberdade e responsabilidade, ou seja, é falar de ética.

Na verdade, quando os nossos antepassados afiavam pedras de sílex para fazer facas, usavam-nas tanto para cortar a pele dos vestuários quanto para se matarem uns aos outros. O mesmo se pode dizer de outras tecnologias muito mais avançadas, como a energia produzida pela fusão de átomos, como ocorre no Sol, que certamente poderia ser utilizada para produzir energia limpa e renovável, mas também para reduzir o nosso planeta a um monte de cinzas.

A inteligência artificial, no entanto, é um instrumento ainda mais complexo. Quase diria que se trata de um instrumento *sui generis*. Assim, enquanto o uso de uma ferramenta simples (como a faca) está sob o controlo do ser humano que a utiliza e o seu bom uso depende somente deste, a inteligência artificial, ao contrário, pode adaptar-se autonomamente à tarefa que lhe é atribuída e, se for projetada dessa forma, fazer escolhas independentes do ser humano para alcançar o objetivo estabelecido[7].

Convém sempre recordar que a máquina pode, sob algumas formas e com estes novos meios, produzir escolhas algorítmicas. O que a máquina faz é uma escolha técnica entre várias possibilidades e baseia-se ou em critérios bem definidos ou em inferências estatísticas. Pelo contrário, o ser humano não só escolhe como, no seu coração, é capaz de decidir. A decisão é um elemento que poderíamos definir como o mais estratégico de uma escolha e requer uma avaliação prática. Frequentemente, na difícil tarefa de governar, somos chamados a tomar decisões com consequências para muitas pessoas. A esse respeito, a reflexão humana sempre falou de sabedoria, a *phronesis* da filosofia grega e, pelo menos em parte, a sabedoria da Sagrada Escritura. Diante dos prodígios das máquinas, que parecem saber escolher de forma independente, devemos ter bem claro que a decisão deve ser sempre deixada ao ser humano, mesmo sob os tons dramáticos e urgentes com que, às vezes, se apresenta na nossa vida. Condenaríamos a humanidade a um futuro sem esperança se retirássemos às pessoas a capacidade de decidir sobre si mesmas e sobre as suas vidas, obrigando-as a depender das escolhas das máquinas. Precisamos de garantir e proteger um espaço de controle significativo do ser humano sobre o processo de escolha dos programas de inteligência artificial: está em jogo a própria dignidade humana.

Permitam-me insistir precisamente sobre este tema: num drama como o dos conflitos armados, é urgente repensar o desenvolvimento e o uso de dispositivos como as chamadas “armas autónomas letais”, a fim de banir a sua utilização, começando desde já pelo compromisso efetivo e concreto de introduzir um controlo humano cada vez mais significativo. Nenhuma máquina, em caso algum, deveria ter a possibilidade de optar por tirar a vida a um ser humano.

Acresce que o bom uso, pelo menos das formas avançadas de inteligência artificial, não estará totalmente sob o controlo nem dos utilizadores nem dos programadores que, no momento da conceção, definiram os seus objetivos originais. E isso é tanto mais verdadeiro quanto é altamente provável que, num futuro não distante, os programas de inteligência artificial possam comunicar diretamente entre si para melhorar o seu desempenho. E se, no passado, os seres humanos que moldaram ferramentas simples viram a sua existência moldada por elas – a faca permitiu-lhes sobreviver ao frio, mas também desenvolver a arte da guerra – agora que moldaram um instrumento complexo, verão este último moldar ainda mais a sua existência[8].

O mecanismo básico da inteligência artificial

Gostaria agora de me deter brevemente sobre a complexidade da inteligência artificial. Na sua essência, a inteligência artificial é uma ferramenta projetada para a resolução de um problema e funciona através de um encadeamento lógico de operações algébricas, realizadas sobre categorias de dados, que são comparados para descobrir correlações, melhorando o seu valor estatístico, graças a um processo de autoaprendizagem, baseado na busca de novos dados e na auto-modificação dos seus procedimentos de cálculo.

A inteligência artificial está assim projetada para resolver problemas específicos, mas, para quem a utiliza, é frequentemente irresistível a tentação de tirar deduções gerais, mesmo antropológicas, a partir das soluções específicas que propõe.

Um bom exemplo é o uso de programas projetados para ajudar os magistrados nas decisões relativas à concessão de prisão domiciliária a reclusos que cumprem pena num estabelecimento prisional. Neste caso, pede-se à inteligência artificial que preveja a probabilidade de reincidência no crime cometido por um condenado a partir de categorias preestabelecidas (tipo de crime, comportamento na prisão, avaliação psicológica, etc.), permitindo que a inteligência artificial tenha acesso a categorias de dados relativos à vida privada do preso (origem étnica, nível de instrução, linha de crédito, etc.). O uso de tal metodologia – que por vezes corre o risco de delegar *de facto* numa máquina a última palavra sobre o destino de uma pessoa – pode implicitamente trazer consigo a referência aos preconceitos inerentes às categorias de dados utilizados pela inteligência artificial.

Estar classificado num determinado grupo étnico ou, mais prosaicamente, ter cometido anos antes uma infração menor (por exemplo, não ter pago uma multa de estacionamento) influenciará a decisão sobre a concessão da prisão domiciliária. Porém, o ser humano está sempre em evolução e é capaz de surpreender com as suas ações, algo que uma máquina não pode ter em consideração.

Deve-se ainda observar que aplicações semelhantes à mencionada sofrerão uma aceleração pelo facto dos programas de inteligência artificial estarem cada vez mais dotados com a capacidade de interagir diretamente com os seres humanos (*chatbots*), mantendo conversas com eles e estabelecendo relações de proximidade, frequentemente, muito agradáveis e reconfortantes, uma vez que esses programas de inteligência artificial serão projetados para aprender a responder, de forma personalizada, às necessidades físicas e psicológicas dos seres humanos.

Esquecer que a inteligência artificial não é outro ser humano e que não pode propor princípios gerais, é muitas vezes um erro grave que decorre ou da profunda necessidade de os seres humanos encontrarem uma forma estável de companhia ou dum pressuposto subconsciente, isto é, do pressuposto de que as observações conseguidas mediante um mecanismo de cálculo sejam dotadas de qualidades de certeza indiscutível e de universalidade inquestionável.

No entanto, este pressuposto é arriscado, como demonstra o exame dos limites intrínsecos do próprio cálculo. A inteligência artificial usa operações algébricas a realizar segundo uma sequência lógica (por exemplo, se o valor de X for superior ao de Y, multiplica X por Y; caso contrário, divide X por Y). Este método de cálculo – o chamado "algoritmo" – não tem nem objetividade nem neutralidade[9]. Sendo baseado na álgebra, só pode examinar realidades formalizadas em termos numéricos[10].

Além disso, não se deve esquecer que os algoritmos projetados para resolver problemas muito complexos são tão sofisticados que é difícil, para os próprios programadores, compreenderem exatamente como conseguem alcançar os seus resultados. Esta tendência para a sofisticação corre o risco de se acelerar consideravelmente com a introdução de computadores quânticos que não funcionarão de acordo com circuitos binários (semicondutores ou microchips) mas segundo as leis, bastante complexas, da física quântica. Por outro lado, a introdução contínua de microchips, cada vez mais eficientes, já se tornou uma das causas da dominância do uso da inteligência artificial por parte das poucas nações que com ela estão equipadas.

Sofisticadas ou não, a qualidade das respostas que os programas de inteligência artificial fornecem depende, em última análise, dos dados que usam e da forma como são estruturados.

Por fim, gostaria de assinalar um último campo, no qual emerge claramente a complexidade do mecanismo da chamada inteligência artificial generativa (*Generative Artificial Intelligence*). Ninguém duvida que existem hoje magníficos instrumentos de acesso ao conhecimento, permitindo até o *self-learning* e o *self-tutoring* numa infinidade de áreas. Muitos de nós ficámos impressionados com as aplicações, facilmente disponíveis online, para redigir um texto ou produzir uma imagem sobre qualquer tema ou assunto. Particularmente fascinados por esta perspetiva são os estudantes, que as utilizam de forma desproporcionada quando precisam de preparar trabalhos escolares.

No entanto, estes alunos, frequentemente muito mais preparados e habituados ao uso da inteligência artificial do que os seus professores, esquecem que a chamada inteligência artificial generativa, em sentido estrito, não é propriamente "generativa". Na verdade, busca nos *big data* informações, elaborando-as segundo o estilo que lhe foi solicitado. Nem desenvolve conceitos nem análises novas. Repete as que encontra, dando-lhes uma forma apelativa. E quanto mais uma noção ou uma hipótese se repete, mais a considera legítima e válida. Em vez de "generativa", ela é "reforçadora", no sentido de que reorganiza os conteúdos existentes, contribuindo para consolidá-los, muitas vezes sem verificar se contêm erros ou preconceitos.

Deste modo, não só se corre o risco de legitimar *fake news* e de reforçar a vantagem de uma cultura dominante, mas igualmente de minar o processo educativo *in nuce*. A educação, que deveria fornecer aos estudantes a possibilidade de uma reflexão autêntica, corre o risco de se reduzir a uma repetição de noções que, cada vez mais, serão consideradas incontestáveis, simplesmente por causa da sua contínua repetição[11].

Colocar novamente a dignidade da pessoa no centro em vista de uma proposta ética comum

Uma observação mais geral deve-se agora acrescentar ao que já foi dito. A era de inovação tecnológica, que estamos atravessando, é acompanhada por uma conjuntura social particular e sem precedentes: sobre os grandes temas da vida social, torna-se cada vez mais difícil encontrar consensos. Mesmo em comunidades caracterizadas por uma certa continuidade cultural, surgem frequentemente debates acesos e confrontos que dificultam a produção de reflexões e soluções políticas comuns orientadas para a procura do que é bom e justo. Além da complexidade de visões legítimas que caracterizam a família humana, surge um fator que parece unir essas diferentes instâncias. Regista-se como que uma perda ou, pelo menos, um eclipse do sentido do humano e uma aparente insignificância do conceito de dignidade humana[12]. Parece que se está a perder o valor e o significado profundo de uma das categorias fundamentais do Ocidente: a categoria de pessoa humana. E assim, nesta era em que os programas de inteligência artificial questionam o ser humano e as suas ações, é precisamente a fraqueza do *ethos* ligado à perceção do valor e da dignidade da pessoa humana que corre o risco de ser o maior *vulnus* na implementação e no desenvolvimento destes sistemas. Efetivamente, não devemos esquecer que nenhuma inovação é neutra. A tecnologia nasce com um propósito e, com o seu impacto na sociedade humana, representa sempre uma forma de ordem nas relações sociais e uma disposição

de poder, permitindo a uns realizar determinadas ações, enquanto a outros impede de concretizar outras. Esta dimensão constitutiva de poder da tecnologia inclui sempre, de uma maneira mais ou menos explícita, a visão do mundo de quem a criou e desenvolveu.

O mesmo se aplica aos programas de inteligência artificial. Para que estes sejam instrumentos de construção do bem e de um amanhã melhor, devem estar sempre orientados ao bem de cada ser humano. Devem ter uma inspiração ética.

Com efeito, a decisão ética é aquela que tem em conta não apenas os resultados de uma ação, mas também os valores em jogo e os deveres que deles derivam. Por isso, congratulei-me com a assinatura da *Rome Call for AI Ethics*[13], na cidade de Roma em 2020, e com o seu apoio a essa forma de moderação ética dos algoritmos e programas de inteligência artificial, que designei de “algor-ética”[14]. Num contexto plural e global, em que se apresentam também sensibilidades diferentes e hierarquias plurais nas escalas de valores, poderia parecer difícil encontrar uma única hierarquia de valores. Mas na análise ética, também podemos recorrer a outros tipos de instrumentos: se tivermos dificuldade em definir um único conjunto de valores globais, podemos encontrar alguns princípios comuns com os quais enfrentar e resolver eventuais dilemas ou conflitos da vida.

Foi por esta razão que nasceu a *Rome Call*: no termo “algor-ética” condensam-se uma série de princípios que se revelam uma plataforma global e plural capaz de encontrar o apoio de culturas, religiões, organizações internacionais e grandes empresas, que são protagonistas deste desenvolvimento.

A política de que precisamos

Porque é inerente ao seu mecanismo fundamental, não podemos esconder o risco concreto de que a inteligência artificial limita a visão do mundo a realidades expressáveis em números e encerradas em categorias pré-concebidas, excluindo o contributo de outras formas de verdade e impondo modelos antropológicos, socioeconómicos e culturais uniformes. O paradigma tecnológico incarnado pela inteligência artificial corre, pois, o risco de dar lugar a um paradigma muito mais perigoso, que identifiquei como “paradigma tecnocrático”[15]. Não podemos permitir que um instrumento tão poderoso e indispensável como a inteligência artificial reforce tal paradigma; pelo contrário, devemos fazer da inteligência artificial precisamente um baluarte contra a sua expansão.

E é exatamente aqui que a ação política é urgente, como lembra a Encíclica *Fratelli tutti*. É certo que «muitos possuem uma má noção da política, e não se pode ignorar que frequentemente, por trás deste facto, estão os erros, a corrupção e a ineficiência de alguns políticos. A isto vêm juntar-se as estratégias que visam enfraquecê-la, substituí-la pela economia ou dominá-la por alguma ideologia. E contudo poderá o mundo funcionar sem política? Poderá encontrar um caminho eficaz para a fraternidade universal e a paz social sem uma boa política?»[16].

A nossa resposta a estas últimas perguntas é: não! A política é necessária! Quero repetir nesta ocasião que «perante tantas formas de política mesquinhas e fixadas no interesse imediato [...], a grandeza política mostra-se quando, em momentos difíceis, se trabalha com base em grandes princípios e pensando no bem comum a longo prazo. O poder político tem muita dificuldade em assumir este dever num projeto de naçãoe, mais ainda, num projeto comum para a humanidade presente e futura»[17].

Prezadas Senhoras, Ilustres Senhores!

Esta minha reflexão sobre os efeitos da inteligência artificial no futuro da humanidade leva-nos assim a considerar a importância duma “política sã” para olharmos o nosso futuro com esperança e confiança. Como já disse noutros lugares, «a sociedade mundial tem graves carências estruturais que não se resolvem com remendos ou soluções rápidas meramente ocasionais. Há coisas que devem ser mudadas com reajustamentos profundos e transformações importantes. E só uma política sã poderia conduzir o processo, envolvendo os mais diversos setores e os conhecimentos mais variados. Desta forma, uma economia integrada num projeto político, social, cultural e popular que vise o bem comum pode “abrir caminho a oportunidades diferentes, que não

implica frenar a criatividade humana nem o seu sonho de progresso, mas orientar esta energia por novos canais" (*Laudato si',*191»[18].

Éprecisamente este o caso da inteligência artificial. Cabe a todos nós fazer um bom uso dela, e cabe à política criar as condições para que essa boa utilização seja possível e frutuosa.

Obrigado.

---

[1] Francisco, *Mensagem para a LVII Jornada Mundial da Paz (1/I/2024),* 1.

[2] Cf. *ibid*.

[3] Cf. *ibid.,* 2.

[4] Esta ambivalência foi já percebida pelo Papa São Paulo VI no seu *Discurso aos funcionários do "Aloysianum de Gallarate"* (19/VI/1964).

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

[6] Cf. Francisco, Carta enc. *Laudato si'* (24/V/2015), 102-114.

[7] Cf. Francisco, *Mensagem para a LVII Jornada Mundial da Paz (1/I/2024)*, 3.

[8] As intuições de Marshall McLuhan e de John M. Culkin são particularmente pertinentes para as consequências do uso da inteligência artificial.

[9] Cf. Francisco, *Discurso aos participantes da Plenária da Pontifícia Academia para a Vida* (28/II/2020).

[10] Cf. Francisco, *Mensagem para a LVII Jornada Mundial da Paz (1/I/2024)*, 4.

[11] Cf. Francisco, *Mensagem para a LVII Jornada Mundial da Paz (1/I/2024)*, 3 e 7.

[12] Cf. Dicastério para a Doutrina da Fé, Declaração *Dignitas infinita* sobre a dignidade humana (2/IV/2024).

[13] Cf. Francisco, *Discurso aos participantes da Plenária da Pontifícia Academia para a Vida* (28/II/2020).

[14] Cf. Francisco, *Discurso aos participantes no Congresso sobre a dignidade dos menores no mundo digital* (14/XI/2019); Cf. *Discurso aos participantes da Plenária da Pontifícia Academia para a Vida* (28/II/2020).

[15] Para uma exposição mais ampla, remeto para a minha Carta Encíclica *Laudato si'* sobre o cuidado da casa comum de 24 de maio de 2015.

[16] Carta Enc. *Fratelli tutti* sobre a fraternidade e a amizade social (3/X/2020), 176

[17] *Ibid*., 178.

[18] *Ibid.,* 179.

[01030-PO.01] [Texto original: Italiano]

Traduzione in lingua polacca

Fascynujące i groźne narzędzie

Szanowne Panie, Dostojni Panowie!

Zwracam się dziś do was, Przywódców Forum Międzyrządowego G7, z refleksją na temat wpływu sztucznej inteligencji na przyszłość ludzkości.

„Pismo Święte zaświadcza, że Bóg dał ludziom swego Ducha, aby posiadali «mądrość, rozum, wiedzę i znajomość wszelkiego rzemiosła» (*Wj* 35, 31)”[1]. Nauka i technologia są zatem niezwykłymi produktami twórczego potencjału nas, istot ludzkich[2].

Zatem, właśnie z wykorzystania tego twórczego potencjału, który dał nam Bóg, rodzi się sztuczna inteligencja.

Jak wiadomo, jest ona narzędziem niezwykle potężnym, wykorzystywanym w bardzo wielu obszarach ludzkiego działania: od medycyny po świat pracy, od kultury po sferę komunikacji, od edukacji po politykę. I można teraz bezpiecznie założyć, że jego wykorzystanie będzie w coraz większym stopniu wpływać na nasz sposób życia, nasze relacje społeczne, a w przyszłości nawet na sposób, w jaki pojmujemy naszą tożsamość jako istoty ludzkiej[3].

Temat sztucznej inteligencji jest jednak często postrzegany jako ambiwalentny: z jednej strony fascynuje możliwościami, jakie oferuje, a z drugiej generuje lęk przed konsekwencjami, jakie zapowiada. W związku z tym można powiedzieć, że w nas wszystkich, choć w różnym wymiarze, przeplatają się dwa uczucia: jesteśmy nastawieni entuzjastycznie, gdy wyobrażamy sobie postęp, jaki można uzyskać dzięki sztucznej inteligencji, ale jednocześnie odczuwamy lęk, gdy widzimy zagrożenia związane z jej użyciem[4].

Zresztą, nie możemy wątpić, że pojawienie się sztucznej inteligencji stanowi prawdziwą rewolucję poznawczo-przemysłową, która przyczyni się do stworzenia nowego systemu społecznego, charakteryzującego się złożonymi przemianami epokowymi. Na przykład, sztuczna inteligencja może pozwolić na demokratyzację dostępu do wiedzy, gwałtowny postęp badań naukowych, możliwość powierzenia uciążliwych prac maszynom; ale jednocześnie może przynieść ze sobą większą niesprawiedliwość między krajami rozwiniętymi a rozwijającymi się, między dominującymi a uciskanymi warstwami społecznymi, zagrażając w ten sposób możliwości „kultury spotkania”, a sprzyjając „kulturze odrzucenia”.

Zakres tych złożonych przemian jest oczywiście powiązany z szybkim rozwojem technologicznym samej sztucznej inteligencji.

To właśnie ten dynamiczny postęp technologiczny sprawia, że sztuczna inteligencja jest *narzędziem*, które jest jednocześnie *fascynujące i groźne* i wymaga refleksji na miarę sytuacji.

Być może moglibyśmy zacząć od spostrzeżenia, że sztuczna inteligencja jest przede wszystkim *narzędziem*. I spontanicznie nasuwa się stwierdzenie, że korzyści lub szkody, jakie przyniesie, będą zależeć od jej zastosowania.

Jest to z pewnością prawda, ponieważ tak było w przypadku każdego narzędzia zbudowanego przez istoty ludzkie od zarania dziejów.

Ta nasza zdolność do konstruowania narzędzi, w ilości i złożoności niespotykanej pośród istot żywych, sprawia, że można mówić o *kondycji techniczno-ludzkiej*: istota ludzka zawsze utrzymywała relację ze środowiskiem,

zapośredniczoną przez narzędzia, jakie stopniowo tworzyła. Nie da się oddzielić historii człowieka i cywilizacji od historii tych narzędzi. Niektórzy chcieli odczytać w tym wszystkim rodzaj braku, niedoboru istoty ludzkiej, tak jakby z powodu tego niedoboru zmuszona była stworzyć technologię[5]. W rzeczywistości, uważne i obiektywne spojrzenie pokazuje nam coś wręcz przeciwnego. Żyjemy w stanie „przekraczania siebie" w odniesieniu do naszej biologicznej istoty; jesteśmy istotami pozbawionymi równowagi w stosunku do tego, co na zewnątrz nas, rzeczywiście, radykalnie otwartymi na to, co poza nami. To stąd bierze początek nasza otwartość na innych i na Boga; to stąd wyrasta twórczy potencjał naszej inteligencji w zakresie kultury i piękna; stąd, w końcu, pochodzi nasza zdolność techniczna. Technologia jest więc śladem tego naszego stanu „przekraczania siebie".

Jednak korzystanie z naszych narzędzi nie zawsze jest jednoznacznie ukierunkowane na dobro. Nawet jeśli istota ludzka czuje w sobie powołanie do przekraczania siebie i do wiedzy przeżywanej jako narzędzie dobra w służbie braciom i siostrom oraz *wspólnemu domowi* (por. *Gaudium et spes*, 16), nie zawsze tak się dzieje. Wręcz przeciwnie, nierzadko, właśnie z powodu swojej radykalnej wolności, ludzkość wypaczyła cel swojego istnienia, stając się wrogiem samej siebie i planety[6]. Ten sam los może spotkać narzędzia technologiczne. Tylko wtedy, gdy zagwarantuje się ich powołanie do służby człowiekowi, narzędzia technologiczne ukażą nie tylko wielkość i wyjątkową godność istoty ludzkiej, ale także zadanie, jaki ona otrzymała, aby „pielęgnować i strzec" (por. *Rdz* 2, 15) planety i wszystkich jej mieszkańców. Mówienie o technologii jest mówieniem o tym, co oznacza bycie człowiekiem, a zatem o naszej wyjątkowej kondycji między wolnością a odpowiedzialnością, to znaczy mówienie o etyce.

Kiedy nasi przodkowie ostrzyli krzemienne kamienie, aby skonstruować noże, używali ich zarówno do cięcia skóry na ubrania, jak i do zabijania się nawzajem. To samo można by powiedzieć o innych, znacznie bardziej zaawansowanych technologiach, takich jak energia wytwarzana przez fuzję atomów, taką jaka zachodzi na Słońcu, która z pewnością mogłaby zostać wykorzystana do produkcji czystej, odnawialnej energii, ale także do sprowadzenia naszej planety do stosu popiołu.

Sztuczna inteligencja jest jednak narzędziem jeszcze bardziej złożonym. Powiedziałbym wręcz, że jest to narzędzie *sui generis.* Tak więc, podczas gdy użycie prostego narzędzia (takiego jak nóż) jest pod kontrolą istoty ludzkiej, która go używa, i tylko od niej zależy jego dobre użycie, to sztuczna inteligencja, przeciwnie, może autonomicznie dostosowywać się do przydzielonego jej zadania i, jeśli jest zaprojektowana w ten sposób, dokonywać wyborów niezależnych od istoty ludzkiej, aby osiągnąć wyznaczony cel[7].

Należy zawsze pamiętać, że maszyna może, w pewnych formach i za pomocą tych nowych środków, dokonywać wyborów algorytmicznych. To, co robi maszyna, jest jednym wyborem technicznym spośród wielu możliwości, i opiera się albo na dobrze zdefiniowanych kryteriach, albo na wnioskach statystycznych. Natomiast istota ludzka nie tylko wybiera, ale jest zdolna do podejmowania decyzji w sercu. Decyzja jest elementem, który moglibyśmy nazwać bardziej strategicznym niż jakiś wybór, i wymaga ona praktycznej oceny. Czasami, często w trudnym zadaniu rządzenia, jesteśmy wezwani do podjęcia decyzji, które będą miały konsekwencje także dla wielu osób. Ludzka refleksja od zawsze mówi w tym względzie o mądrości, o *phronesis* filozofii greckiej i przynajmniej częściowo o mądrości Pisma Świętego. W obliczu cudów maszyn, które zdają się być zdolne do samodzielnego wyboru, musimy sobie wyraźnie uświadomić, że istocie ludzkiej zawsze należy pozostawić decyzję, nawet z odcieniem dramatycznym i naglącym, jakie czasami są obecne w naszym życiu. Skazalibyśmy ludzkość na przyszłość bez nadziei, gdybyśmy pozbawili ludzi zdolności do decydowania o sobie i swoim życiu, skazując ich na zależność od wyborów dokonywanych przez maszyny. Potrzebujemy gwarancji i ochrony pewnej przestrzeni znaczącej kontroli dokonywanej przez istoty ludzkie, nad procesem wyboru programów sztucznej inteligencji: chodzi w tym o samą godność ludzką.

Właśnie w tej kwestii pozwólcie mi, że będę nalegał: w dramacie, jakim jest konflikt zbrojny, należy pilnie przemyśleć rozwój i wykorzystanie urządzeń takich jak tak zwana „śmiercionośna broń autonomiczna", aby zakazać ich używania, począwszy już od faktycznego i konkretnego zobowiązania do wprowadzenia coraz bardziej znaczącej ludzkiej kontroli. Żadna maszyna nie powinna nigdy decydować, czy odebrać życie istocie ludzkiej.

Co więcej, należy dodać, że właściwe wykorzystanie, przynajmniej zaawansowanych form sztucznej inteligencji,

nie będzie w pełni kontrolowane ani przez użytkowników, ani przez programistów, którzy zdefiniowali ich pierwotny cel w chwili ich opracowania. Tym bardziej, że jest wysoce prawdopodobne, iż w niedalekiej przyszłości programy sztucznej inteligencji będą w stanie komunikować się bezpośrednio ze sobą w celu udoskonalenia swoich *osiągnięć*. I jeśli w przeszłości ludzie, którzy projektowali proste narzędzia, widzieli, jak kształtują one ich egzystencję – nóż pozwolił im przetrwać zimno, ale także rozwinąć sztukę wojenną – teraz, gdy ludzie zaprojektowali złożone narzędzie, zobaczą, jak to ostatnie jeszcze bardziej kształtuje ich egzystencję[8].

Podstawowy mechanizm sztucznej inteligencji

Chciałbym teraz zatrzymać się krótko nad złożonością sztucznej inteligencji. W swojej istocie, sztuczna inteligencja jest narzędziem zaprojektowanym do rozwiązywania problemów i funkcjonuje za pomocą logicznego łańcucha operacji algebraicznych, przeprowadzanych na kategoriach danych, które są porównywane w celu odkrycia korelacji, poprawiając ich wartość statystyczną, dzięki procesowi samodzielnego uczenia się, opartemu na poszukiwaniu dalszych danych i samodzielnej modyfikacji jego procedur obliczeniowych.

Sztuczna inteligencja jest zatem zaprojektowana do rozwiązywania specyficznych problemów, jednak u tych, którzy jej używają, pojawia się często nieodparta pokusa wyciągania ogólnych wniosków z konkretnych rozwiązań, jakie proponuje, nawet natury antropologicznej.

Dobrym przykładem jest wykorzystanie programów zaprojektowanych aby pomagać sędziom w podjęciu decyzji o przyznaniu aresztu domowego więźniom odbywającym karę w zakładzie karnym. W tym przypadku wymaga się od sztucznej inteligencji przewidywania prawdopodobieństwa ponownego popełnienia przestępstwa przez skazanego, na podstawie wcześniej określonych kategorii (rodzaj przestępstwa, zachowanie w więzieniu, ocena psychologiczna i inne), pozwalając sztucznej inteligencji na dostęp do kategorii danych związanych z życiem prywatnym skazanego (pochodzenie etniczne, poziom wykształcenia, linia kredytowa i inne). Korzystanie z takiej metodologii – która czasami grozi przekazaniem *de facto* maszynie ostatniego słowa w sprawie losu danej osoby – może domyślnie nieść ze sobą odniesienie do uprzedzeń, nieodłącznie związanych z kategoriami danych wykorzystywanych przez sztuczną inteligencję.

Bycie zaklasyfikowanym do określonej grupy etnicznej lub, bardziej prozaicznie, popełnienie drobnego przestępstwa wiele lat wcześniej (niezapłacenie, na przykład, mandatu za nieprawidłowe parkowanie), w rzeczywistości wpłynie na decyzję o przyznaniu aresztu domowego. Jednakże istota ludzka cały czas ewoluuje i potrafi zaskakiwać swoimi działaniami, czego maszyna nie jest w stanie wziąć pod uwagę.

Należy również zauważyć, że zastosowania podobne do właśnie wspomnianego, zostaną przyspieszone przez fakt, że programy sztucznej inteligencji będą w coraz większym stopniu wyposażone w zdolność do bezpośredniej interakcji z ludźmi (*chatboty*), prowadząc z nimi rozmowy i nawiązując z nimi relacje bliskości, często bardzo przyjemne i kojące, ponieważ te programy sztucznej inteligencji zostaną zaprojektowane tak, aby nauczyć się odpowiadać w formie spersonalizowanej na fizyczne i psychologiczne potrzeby istot ludzkich.

Zapominanie, że sztuczna inteligencja nie jest inną istotą ludzką, i że ona nie może zaproponować ogólnych zasad, jest często poważnym błędem, który wynika albo z głębokiej potrzeby istot ludzkich do znalezienia stabilnej formy towarzystwa, albo z ich podświadomego założenia, czyli założenia, że obserwacje uzyskane za pomocą mechanizmu obliczeniowego są obdarzone cechami niekwestionowanej pewności i niewątpliwej uniwersalności.

To założenie jest jednak ryzykowne, jak pokazuje analiza nieodłącznych ograniczeń samych obliczeń. Sztuczna inteligencja wykorzystuje operacje algebraiczne, które mają być wykonywane zgodnie z logiczną sekwencją (np. jeśli wartość X jest większa niż wartość Y, pomnóż X przez Y; w przeciwnym razie podziel X przez Y). Ta metoda obliczeń – tak zwany „algorytm” – nie jest ani obiektywna, ani neutralna[9]. Opierając się na algebrze, może jedynie badać rzeczywistość sformalizowaną w kategoriach liczbowych[10].

Co więcej, nie należy zapominać, że algorytmy zaprojektowane do rozwiązywania bardzo złożonych problemów

są tak zaawansowane, że samym programistom sprawia trudność dokładne zrozumienie, w jaki sposób osiągają swoje wyniki. Ta tendencja do wyrafinowania grodzi znacznym przyspieszeniem wraz z wprowadzeniem komputerów kwantowych, które nie będą działać w oparciu o obwody binarne (półprzewodniki lub mikroprocesory), ale według dość złożonych praw fizyki kwantowej. Z drugiej strony, ciągłe wprowadzanie coraz bardziej wydajnych mikroprocesorów już stało się jedną z przyczyn dominacji wykorzystania sztucznej inteligencji przez kilka krajów, które są w nią wyposażone.

Niezależnie od tego, czy są one wyrafinowane, czy nie, jakość odpowiedzi udzielanych przez programy sztucznej inteligencji ostatecznie zależy od danych, z których one korzystają, i od tego, jak są ustrukturyzowane.

Na koniec pozwolę sobie wskazać jeszcze jeden obszar, w którym wyraźnie ujawnia się złożoność mechanizmu tak zwanej generatywnej sztucznej inteligencji (*Generative Artificial Intelligence*). Nikt nie ma wątpliwości, że obecnie dostępne są wspaniałe narzędzia dostępu do wiedzy, które pozwalają nawet na *samouczenie się* i *samokształcenie* w niezliczonych dziedzinach. Wielu z nas było pod wrażeniem aplikacji, łatwo dostępnych online, do komponowania tekstu lub tworzenia obrazu na dowolny temat. Szczególnie przyciągają one uwagę uczniów, którzy, gdy muszą przygotować wypracowania, korzystają z nich zbyt często.

Ci uczniowie, którzy często są znacznie lepiej przygotowani i przyzwyczajeni do korzystania ze sztucznej inteligencji niż ich profesorowie, zapominają jednak, że tak zwana generatywna sztuczna inteligencja, ściśle mówiąc, nie jest tak naprawdę „generatywna". W rzeczywistości, przeszukuje ona *big data* w poszukiwaniu informacji i przetwarza je w żądanym od niej stylu. Nie opracowuje nowych koncepcji ani analiz. Powtarza te, które znajdzie, nadając im atrakcyjną formę. A im częściej znajduje powtarzające się pojęcie lub hipotezę, tym bardziej uważa je za uzasadnione i ważne. Jest ona zatem nie tyle „generatywna" ile „utrwalająca", w tym sensie, że porządkuje istniejące treści, pomagając je skonsolidować, często nie sprawdzając, czy zawierają błędy lub z góry przyjęte koncepcje.

W ten sposób nie tylko pojawia się ryzyko legitymizacji fałszywych wiadomości (*fake news*) i wzmocnienia przewagi jakiejś dominującej kultury, ale także podważa to w istocie proces edukacyjny. Edukacja, która powinna zapewnić uczniom możliwość autentycznej refleksji, może zostać sprowadzona do powtarzania pojęć, które będą coraz częściej oceniane jako niebudzące zastrzeżeń, jedynie z powodu ich ciągłego powtarzania[11].

Postawienie w centrum godności osoby w perspektywie wspólnej propozycji etycznej

Do tego, co zostało już powiedziane, należy teraz dodać spostrzeżenie bardziej ogólne. Okres innowacji technologicznych, przez który obecnie przechodzimy, idzie w parze ze szczególną i bezprecedensową sytuacją społeczną: coraz trudniej znaleźć porozumienie w głównych kwestiach życia społecznego. Nawet w społecznościach charakteryzujących się pewną ciągłością kulturową, często dochodzi do gorących debat i konfrontacji, co utrudnia tworzenie wspólnych refleksji i rozwiązań politycznych, mających na celu poszukiwanie tego co dobre i sprawiedliwe. Poza złożonością słusznych wizji, które charakteryzują ludzką rodzinę, pojawia się czynnik, który wydaje się łączyć te różne przypadki. Mamy do czynienia z utratą, lub przynajmniej przysłonięciem, poczucia człowieczeństwa i pozorną nieistotnością pojęcia godności ludzkiej[12]. Wydaje się, że zatraca się wartość i głębokie znaczenie jednej z podstawowych kategorii Zachodu: kategorii osoby ludzkiej. I oto w tym okresie, w którym programy sztucznej inteligencji poddają w wątpliwość istotę ludzką i jej działanie, to właśnie słabość *etosu* związanego z postrzeganiem wartości i godności osoby ludzkiej, może być największym *vulnus* (słabością) we wdrażaniu i rozwoju tych systemów. Nie możemy zapominać, że żadna innowacja nie jest neutralna. Technologia rodzi się w określonym celu i jej wpływ na społeczeństwo ludzkie zawsze reprezentuje pewną formę ładu w relacjach społecznych i układ władzy, umożliwiając niektórym wykonywanie działań i uniemożliwiając innym wypełnianie innych czynności. Ten konstytutywny wymiar władzy technologii zawsze obejmuje, w mniej lub bardziej wyraźny sposób, światopogląd tych, którzy ją zrealizowali i wypracowali.

Dotyczy to również programów sztucznej inteligencji. Aby były one narzędziami budowania dobra i lepszego jutra, muszą być zawsze ukierunkowane na dobro każdego człowieka. Muszą być inspirowane etycznie.

Etyczna decyzja to taka, która bierze pod uwagę nie tylko skutki danego działania, ale także wchodzące w grę

wartości, i obowiązki, które z tych wartości wynikają. Dlatego z zadowoleniem przyjąłem podpisanie *Rzymskiego apelu o etykę AI* w 2020 r.[13] i jego poparcie dla tej formy etycznego moderowania algorytmów i programów sztucznej inteligencji, które nazwałem „algoretyką”[14]. W pluralistycznym i globalnym kontekście, w którym widoczne są również różne wrażliwości i wielorakie hierarchie w skalach wartości, znalezienie jednej hierarchii wartości wydaje się trudne. Jednak w analizie etycznej możemy również uciec się do innych rodzajów narzędzi: jeśli mamy trudności ze zdefiniowaniem jednego zestawu globalnych wartości, możemy jednak znaleźć wspólne zasady, za pomocą których można rozwiązać wszelkie dylematy lub konflikty życia.

Z tego powodu zrodził się *Rzymski apel*: w terminie „algoretyka” skupia się zestaw zasad, które okazują się globalną i pluralistyczną platformą, zdolną do znalezienia wsparcia ze strony kultur, religii, organizacji międzynarodowych i dużych koncernów, kluczowych uczestników tego rozwoju.

Polityka, jakiej potrzebujemy

Nie możemy zatem ukrywać konkretnego zagrożenia, ponieważ jest ono nieodłączne od jego podstawowego mechanizmu, mianowicie, że sztuczna inteligencja ograniczy wizję świata do rzeczywistości wyrażalnej w liczbach i zamkniętej w gotowych kategoriach, wypierając wkład innych form prawdy i narzucając jednolite modele antropologiczne, społeczno-ekonomiczne i kulturowe. Paradygmat technologiczny, którego ucieleśnieniem jest sztuczna inteligencja, niesie ze sobą ryzyko ustąpienia miejsca znacznie bardziej niebezpiecznemu paradygmatowi, który zidentyfikowałem już jako „paradygmat technokratyczny”[15]. Nie możemy pozwolić, aby narzędzie tak potężne i niezbędne jak sztuczna inteligencja, wzmocniło ten paradygmat; wręcz przeciwnie, musimy uczynić sztuczną inteligencję bastionem właśnie przeciwko jego ekspansji.

I to właśnie tutaj pilnie potrzebne są działania polityczne, jak przypomina encyklika *Fratelli tutti*. Z pewnością „dla wielu polityka jest dziś brzydkim słowem i nie można ignorować, że często za tym faktem kryją się błędy, korupcja i nieskuteczność niektórych polityków. Dochodzą do tego strategie, które mają na celu jej osłabienie, zastąpienie jej gospodarką lub zdominowanie jej jakąś ideologią. Ale czy jednak świat może funkcjonować bez polityki? Czy może istnieć skuteczna droga do powszechnego braterstwa i pokoju społecznego bez dobrej polityki?”[16].

Nasza odpowiedź na te ostatnie pytania brzmi: nie! Polityka jest potrzebna! Chcę przy tej okazji powtórzyć, że „w obliczu tak wielu małostkowych i doraźnych form polityki [...] wielkość polityczna ukazuje się wtedy, gdy w trudnych chwilach działamy w oparciu o wielkie zasady i myślimy długofalowo o dobru wspólnym. Władzy politycznej bardzo trudno przyjąć ten obowiązek w planie narodowym, a tym bardziej we wspólnym projekcie dla obecnej i przyszłej ludzkości”[17].

Szanowne Panie, Dostojni Panowie!

Niniejsza refleksja na temat wpływu sztucznej inteligencji na przyszłość ludzkości, prowadzi nas zatem do rozważenia znaczenia „zdrowej polityki”, aby patrzeć w naszą przyszłość z nadzieją i ufnością. Jak już powiedziałem w innym miejscu, „globalne społeczeństwo ma poważne braki strukturalne, których nie można rozwiązywać za pomocą łatek lub szybkich, czysto okazjonalnych napraw. Są kwestie, które trzeba zmienić poprzez gruntowne przemyślenia i poważne transformacje. Może do tego doprowadzić jedynie zdrowa polityka, angażując najbardziej zróżnicowane sektory i najbardziej różnorodne dziedziny wiedzy. W ten sposób gospodarka zintegrowana w ramach projektu politycznego, społecznego, kulturalnego i ludowego, dążącego do dobra wspólnego, może «otworzyć nowe drogi dla różnych możliwości, które nie oznaczają powstrzymywania kreatywności człowieka i jego marzeń o postępie, a raczej ukierunkowują tę energię w nowy sposób» (*Laudato si'*, 191)”[18].

Tak jest właśnie w przypadku sztucznej inteligencji. Obowiązkiem każdego jest jej dobre wykorzystanie, a do obowiązków polityki należy stworzenie warunków, w których takie dobre wykorzystanie będzie możliwe i owocne.

Dziękuję.

[1] *Orędzie na 57. Światowy Dzień Pokoju*, *1 stycznia 2024*, 1.

[2] Por. *tamże.*

[3] Por. tamże, 2.

[4] Tę ambiwalencję dostrzegał już papież Paweł VI, w swoim przemówieniu do pracowników *Ośrodka Automatyzacji Analizy Językowej Aloysianum*, 19 czerwca 1964.

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

[6] Enc. *Laudato si'* (24 maja 2015), 102-114.

[7] Por. *Orędzie na 57 Światowy Dzień Pokoju*, *1 stycznia 2024*, 3.

[8] Szczególnie istotne dla konsekwencji korzystania ze sztucznej inteligencji są spostrzeżenia Marshalla McLuhana i Johna M. Culkina.

[9] Por. *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[10] Por. *Orędzie na 57. Światowy Dzień Pokoju*, *1 stycznia 2024*, 4.

[11] Por. *tamże*, 3 i 7.

[12] Por. Dykasteria Nauki Wiary, Deklaracja *Dignitas infinita* o godności człowieka (2 kwietnia 2024).

[13] Por. *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[14] Por. *Discorso ai partecipanti al Convegno "Promoting Digital Child Dignity – From Concet to Action"*, 14 novembre 2019; *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita,* 28 febbraio 2020.

[15] Aby uzyskać szersze omówienie, odsyłam do mojej encykliki *Laudato si'* o trosce o wspólny dom z 24 maja 2015 r.

[16] Enc. *Fratelli tutti*, O braterstwie i przyjaźni społecznej (3 października 2020), 176.

[17] *Tamże*, 178.

[18] *Tamże*, 179.

[01030-PL.01] [Testo originale: Italiano]

Traduzione in lingua araba

كلمة قداسة البابا فرنسيس

الى مجموعة الدّول الصّناعيّة السّبع (G7)

أداة رائعة ومُرعبة

السّيّدات الأعزّاء، السّادة الكِرام!

أتوجّه اليوم إليكم، أنتم قادة المنتدى الحكوميّ الدّوليّ لمجموعة الدّول الصّناعيّة السّبع، وأفكّر في تأثير الذّكاء الاصطناعيّ على مستقبل البشريّة.

"يقول الكتاب المقدّس أنّ الله أعطى الإنسان روحه ليملأه "مَهارةً وفَهْمًا ومَعرفةً بِجَميعِ الصّنائع" (خروج 35، 31)"[1]. العِلم والتّكنولوجيا هما نتاجان فائقان لقدرة الإنسان الإبداعيّة[2].

ومن استخدامنا لهذه الإمكانات الإبداعيّة التي وهبها الله لنا، وُلِدَ الذّكاء الاصطناعيّ.

وهذا الذّكاء، كما هو معروف، هو أداة قديرة جدًّا، تُستخدم في مجالات عديدة من مجالات العمل الإنسانيّ: في الطّب وعالم العمل، والثّقافة ومجال الاتّصال، والتّربية والسّياسة. ويجوز لنا الآن أن نفترض أنّ استخدامه سيزداد أثره دائمًا في أسلوب حياتنا، وفي علاقاتنا الاجتماعيّة، وسيؤثّر في المستقبل حتّى على الطّريقة التي بها نُدرك هويّتنا ككائنات بشريّة[3].

لموضوع الذّكاء الاصطناعيّ نتيجتان: من جهة، يثير حماسنا بسبب الإمكانيّات التي يقدّمها، ومن جهة أخرى يُولِّد فينا المخاوف بسبب العواقب التي يمكن أن يؤدّي إليها. ويمكن أن نقول إنّنا كلّنا نشعر بهذَين الشّعورَين، ولو بدرجات مختلفة: نحن متحمّسون عندما نتصوّر التّقدّم الذي يمكن أن يأتينا مع الذّكاء الاصطناعيّ، ولكن، في الوقت نفسه، نحن خائفون عندما نرى المخاطر المرتبطة باستخدامه[4].

علاوة على ذلك، لا يمكننا أن نشكّ في أنّ مجيء الذّكاء الاصطناعيّ يمثّل ثورة صناعيّة حقيقيّة في المعرفة، التي ستُسهم في خلق نظام اجتماعيّ جديد يتّسم بتحوّلات تاريخيّة معقدة. على سبيل المثال، يمكن للذّكاء الاصطناعيّ أن يسمح بنوع من الدّيمقراطيّة في الوصول إلى المعرفة، وفي التّقدّم المتسارع للبحث العلميّ، وإمكانيّة تفويض الأعمال الشّاقّة للآلات، وفي الوقت نفسه، يمكن أن يحمل معه ظُلمًا أكبر لبعض الدّول النّامية بالنّسبة للدّول المتقدّمة، وبين الطّبقات الاجتماعيّة المُسيطرة والطّبقات الاجتماعيّة المُظلومة، فيعرّض للخطر إمكانيّة "ثقافة اللقاء" بتأييده لـ"ثقافة الإقصاء".

من الواضح أنّ قدرة هذه التّحوّلات المتعدّدة والمعقدة مرتبطة بالتّطوّر التكنولوجيّ السّريع للذّكاء الاصطناعيّ نفسه.

هذا التّقدّم التكنولوجيّ القويّ نفسه، يجعل من الذّكاء الاصطناعيّ أداة جذّابة ومُخيفة في الوقت نفسه، ويقتضي منّا أن نرتفع بتفكيرنا على المستوى المطلوب.

في هذا المجال، يمكن أن نبدأ من هذه الملاحظة أن الذّكاء الاصطناعيّ هو أوّلًا أداة. ومن الطّبيعي أن نؤكّد أنّ الفوائد أو الأضرار التي سيحملها إلينا تعتمد على استخدامنا له.

هذا صحيح بالتّأكيد، لأنّه هكذا كان الأمر بالنّسبة لكلّ أداة صنعها الإنسان منذ فجر التّاريخ.

قدرتنا هذه على بناء الأدوات، بكمّيّات وإمكانيّات معقدة لا مثيل لها بين الكائنات الحيّة، تجعلنا نتكلّم على ارتباط عميق

لكن استخدامنا لأدواتنا ليس دائمًا موجّهًا إلى الخير. حتّى وإن كان الإنسان يشعر في داخله بأنّه مدعُوٌّ إلى ما هو أبعد منه، وإلى المعرفة التي يستخدمها كأداة لعمل الخير في خدمة الإخوة والأخوات والبيت المشترك (راجع *فرح ورجاء*، 16)، فإنّه لا يتخذ في عمله دائمًا هذا التوجّه إلى الخير. بل، ليس من النّادر أنّ الإنسانيّة ضلّت أهداف وجودها، بفضل نفس حرّيّتها الأساسيّة فيها، تحوّلت إلى عدوّة لنفسها وللكوكب[6]. يمكن أن يكون للأدوات التّكنولوجيّة المصير نفسه. ستبيّن الأدوات التّكنولوجيّة عظمة الإنسان وكرامته الفريدة، بل أيضًا التّفويض الذي كلّفه به الله "ليفلح ويحرس" (راجع تكوين 2، 15) الكوكب وكلّ سكّانه، إذا توفّرت ضمانات أكيدة لإبقاء الأدوات التّكنولوجيّة في خدمة الإنسان. الكلام على التّكنولوجيا يعني الكلام على معنى أن نكون بشرًا، وبالتّالي على حالتنا الفريدة بين الحرّيّة والمسؤوليّة، أي الكلام على الأخلاق.

في الواقع، عندما كان أجدادنا يصقلون حجر الصّوان ليصنعوا السّكاكين، كانوا يستخدمونها لقطع الجلود وصنع الملابس، وكذلك لقتل بعضهم البعض. يمكننا أن نقول الأمر نفسه عن التّقنيّات الأخرى المتطوّرة، مثل الطّاقة النّاتجة عن دمج الذّرات كما يحدث في الشّمس، والتي يمكننا أن نستخدمها بالتّأكيد لإنتاج طاقة نظيفة ومتجدّدة وأيضًا لنحوّل كوكبنا إلى كومة رماد.

مع ذلك، فإن الذّكاء الاصطناعيّ أداة أكثر تعقيدًا. يمكن القول إنّه أداة فريدة من نوعها. وهكذا، في حين أنّ استخدام أداة بسيطة (مثل السّكين) يكون تحت سيطرة الإنسان الذي يستخدمها، وصلاح الاستخدام مرتبط بالإنسان فقط، فإنّ الذّكاء الاصطناعيّ، يمكن أن يتكيّف بشكل مستقلّ بحسب المهمّة التي تحدّد له، وهو يقوم باختيارات مستقلّة عن الإنسان ليصل إلى الهدف المحدّد له، إن تمّ تصميمه بهذه الطّريقة[7].

من الجدير بالذّكر أنّ الآلة يمكنها، في بعض الصّور وبواسطة الوسائل الجديدة، أن تُنتج خيارات خوارزميّة. ما تقوم به الآلة هو عبارة عن اختيار تقنيّ بين احتمالات عديدة، ويرتبط اختيارها إمّا بمعايير محدّدة جيّدًا أو بإملاءات إحصائيّة. وأمّا الإنسان، فإنّه لا يختار فقط، بل هو قادر في قلبه على اتّخاذ القرار. ويمكن أن نقول إنّ القرار أكثر من الاختيار، وهو عنصر استراتيجيّ يتطلّب منّا تقييمًا عمليًّا. أحيانًا، وغالبًا في مهمّة الحكم الصّعبة، نحن مدعوّون إلى أن نتّخذ قرارات لها نتائج، أيضًا فيما يخصّ أشخاصًا كثيرين. في هذا الصّدد، تكلّم التّفكير البشريّ دائمًا على الحكمة، وعلى حكمة الفلسفة اليونانيّة (*phronesis*)، وعلى حكمة الكتاب المقدّس، ولو جزئيًّا. أمام معجزات الآلات، التي تبدو لنا أنّها تعرف أن تختار بشكل مستقلّ، يجب أن يكون واضحًا لنا أنّ القرار يجب أن يبقى دائمًا للإنسان، حتّى في الظّروف المأسويّة والمُلِحّة التي يجب اتخاذ القرار فيها. إن سلبنا الأشخاص قدرتهم على اتّخاذ القرار بشأن أنفسهم وبشأن حياتهم، وحكمنا عليهم بأن يكونوا خاضعين لاختيارات الآلات، فإنّنا نحكم على البشريّة بمستقبل لا رجاءَ فيه. يجب أن نضمن ونحمي مساحة كبيرة لسيطرة الإنسان على عمليّة اختيار برامج الذّكاء الاصطناعيّ: فالكرامة الإنسانيّة نفسها مرهونة بذلك.

حول هذا الموضوع، اسمحوا لِي بأن أؤكّد ما يلي: في مأساة مثل مأساة الصّراعات المسلّحة، من المُلحّ أن نفكّر من جديد في تطوير واستخدام أجهزة مثل التي تُسمّى "الأسلحة القتالة المستقلّة " لكي نحظر استخدامها، ونبدأ بالتزام عمليّ وملموس لكي نُدخل سيطرة الإنسان المهمّة بشكل أكبر دائمًا. لا يجوز على الإطلاق أن يُترك القرار لأيّ آلة في القضاء على حياة إنسان.

علاوة على ذلك، علينا أن نضيف أنّ الاستخدام الجيّد، وعلى الأقلّ للأشكال المتقدّمة من الذّكاء الاصطناعيّ، لن يكون تحت السّيطرة الكاملة، لا تحت سيطرة المستخدمين ولا المبرمجين الذين حدّدوا أهدافها الأصليّة في لحظة البرمجة. هذا صحيح، وصحيح أيضًا ومحتمل جدًّا، أنّ برامج الذّكاء الاصطناعيّ ستتمكّن في مستقبل غير بعيد من أن تتواصل مباشرة بعضها مع بعض لكي تحسّن أدائها. الإنسان القديم الذي صنع أدوات بسيطة في الماضي، رأى أنّ حياته صارت مرتبطة بها، - فالسّكين الذي صنعه سمح له بأن يبقى على قيد الحياة في البرد، وأيضًا بأن يطوّر فنّ الحرب –

الآليّة الأساسيّة للذّكاء الاصطناعيّ

أودّ الآن أن أتوقّف قليلًا عند مدى تعقيد الذّكاء الاصطناعيّ. الذّكاء الاصطناعيّ في أساسه هو أداة مصمّمة من أجل حلّ مشكلة، وتعمل عن طريق سلسلة منطقيّة من العمليّات الجبريّة، التي تتمّ على فئات من البيانات، وتتمّ مقارنتها لاكتشاف الارتباطات فيما بينها، وتحسين قيمتها الإحصائيّة، بفضل تعلّم ذاتيّ، قائم على البحث عن مزيد من البيانات وعلى التّعديل الذّاتيّ لعملياته الحسابيّة.

إذًا، الذّكاء الاصطناعيّ مصمّم لحلّ مشاكل محدّدة، لكن، الذين يستخدمونه معرّضون غالبًا لتجربة لا تقاوم، وهي أن يستخلصوا من الحلول الدّقيقة المحدّدة له استنتاجات عامّة، وحتى ذات طبيعة أنثروبولوجيّة.

مثال على ذلك، استخدام البرامج المصمّمة لمساعدة القُضاة في اتّخاذ القرارات المتعلّقة بالإقامة القسريّة للسّجناء الذين يقضون عقوباتهم في مؤسسة السّجون. في هذه الحالة، يطلبون من الذّكاء الاصطناعيّ أن يتنبّأ باحتماليّة وقوع المحكوم عليه في الجرم ثانية، وذلك بناءً على فئات من البيانات المحدّدة مسبقًا (نوع الجريمة، والسّلوك في السجن، والتّقييم النّفسيّ وغيرها)، ويُسمَح للذّكاء الاصطناعيّ بالوصول إلى فئات من البيانات المتعلّقة بالحياة الخاصّة للسّجين (الأصل العرقيّ، والمستوى التعليميّ، وبيئة الانتماء وغير ذلك). إنّ استخدام مثل هذه المنهجيّة - التي تضع بين يدي الآلة، بحكم الأمر الواقع، الكلمة الأخيرة حول مصير الشّخص - يمكن أن يتضمّن أيضًا إمكانيّة التأثّر بأحكام سابقة متأصّلة في فئات البيانات التي يستخدمها الذّكاء الاصطناعيّ.

الإنسان المصنّف في مجموعة عرقيّة معيّنة، أو مَثَلٌ أبسط، ارتكاب مخالفة صغيرة قبل سنوات (مثلًا، عدم دفع المخالفة للوقوف في مكانٍ ممنوع)، سيؤثّر في الواقع على القرار المتعلّق في تحديد الإقامة القسريّة. لأنّ الإنسان في تطوّر دائم، وهو قادر أن يفاجئ بأفعاله، وهذا أمرٌ لا تستطيع الآلة أن تأخذه بعين الاعتبار.

يجب أن نعرف أيضًا أنّ التّطبيقات الشّبيهة للحالات المذكورة سابقًا، ستتطوّر بسرعة لأنّ برامج الذّكاء الاصطناعيّ ستكون مجهّزة دائمًا بقدرة أكبر على التّفاعل المباشر مع الكائن البشريّ (chatbots)، فتُجري محادثات معهم وتُقيم معهم علاقات وثيقة، غالبًا مُمتعة ومطمئنة جدًّا، إذ سيتمّ تصميم برامج الذّكاء الاصطناعيّ هذه لكي تتعلّم أن تُجيب، وبشكلٍ شخصيّ، على احتياجات البشر الجسديّة والنّفسيّة.

أن ننَسى أنّ الذّكاء الاصطناعيّ ليس إنسانًا آخر، وأنّه لا يستطيع أن يقترح مبادئ عامّة، هذا غالبًا خطأ جسيم ينجم عن حاجة الإنسان العميقة لأن يجد شكلًا ثابتًا من المرافقة، أو عن بديل لها افتراضيّ ولاواعيّ، أو عن الافتراض أنّ البيانات التي يمكن الحصول عليها من الآلآت الحاسبة، لها صفات حقيقيّة لا جدال فيها وهي شاملة ولا شكّ فيها.

مع ذلك، هذا الافتراض فيه مجازفة، كما يبيّن ذلك فحص الحدود الدّاخليّة للحساب نفسه. الذّكاء الاصطناعيّ يستخدم عمليّات جبريّة تتمّ بحسب عمليات منطقيّة (مثلًا، إذا كانت قيمة X أكبر من قيمة Y، فيجب ضرب X في Y، وإلّا فيجب قسمة X على Y). طريقة الحساب هذه – التي تسمّى "الخوارزميّة" - لا تتمتّع لا بالموضوعيّة ولا بالحِيَاد[9]. في الواقع، كونه يقوم على الجَبر، فهو لا يمكنه أن يفحص إلّا الحقائق التي صِيغَت بطريقة رقميّة[10].

علاوة على ذلك، يجب ألّا ننسى أنّ الخوارزميّات المصمّمة لحلّ المشكلات المعقّدة كثيرًا، هي حديثة ومتطوّرة جدًّا، بحيث يصعب على المبرمجين أنفسهم أن يفهموا بالضّبط كيف يمكن لهذه الخوارزميّات أن تحقّق نتائجها. هذا الاتّجاه نحو التّطوّر يهدّد بالتّسارع بشكل كبير وأن يُدخل أجهزة كمبيوتر كموميّة التي لن تعمل مع الدّوائر الثنائيّة (أشباه المُوصِلَات أو الرّقائق الدّقيقة)، بل بحسب قوانين فيزياء الكمّ المعقّدة إلى حدٍّ ما. من ناحية أخرى، الإدخال المستمرّ للرّقائق الدّقيقة والعالية الأداء دائمًا صار أحد أسباب هيمنة استخدام الذّكاء الاصطناعيّ من قِبَل الدّول القليلة التي

سواء كانت متطوّرة أم لا، فإنّ جَودة الإجابات التي تقدّمها برامج الذّكاء الاصطناعيّ تعتمد في نهاية الأمر على البيانات التي تستخدمها وكيفيّة تنظيمها لها.

أخيرًا، أسمح لنفسي أن أُشير إلى مجال أخير فيه يظهر بوضوح تعقيد آليّة ما يسمّى بالذّكاء الاصطناعيّ التّوليديّ (*Generative Artificial Intelligence*). لا أحد يشكّ أنّه يوجد اليوم أدوات رائعة وهي في متناول اليد للوصول إلى المعرفة، التي تسمح لنا حتّى بالتّعلّم الذّاتي والتّثقيف الذّاتي في مجالات كثيرة. أُعجب الكثيرون منّا بالتّطبيقات المتاحة بسهولة عبر الإنترنت لإنشاء نصّ أو إنتاج صورة حول أيّ موضوع أو شخص. انجذب بشكل خاص إلى هذا الجانب الطّلاب الذين يستخدمونه وبصورة مبالغ فيها، عندما يجب عليهم أن يحضّروا عملًا كتابيًّا.

مع ذلك، فإنّ هؤلاء الطّلاب، الذين لهم مهارة ومقدرة على استخدام الذّكاء الاصطناعيّ أكثر من معلّميهم، ينسون أنّ ما يسمّى بالذّكاء الاصطناعيّ التّوليديّ، بالمعنى الضّيّق للكلمة، ليس "توليديًّا" تمامًا. في الحقيقة، هذا الأخير، يبحث عن المعلومات في البيانات الضّخمة ويصنّفها بالأسلوب الذي طُلب منه. ولا يطوّر مفاهيم أو تحاليل جديدة. بل يكرّر الذي يجده ويعطيه شكلًا جذّابًا. وكلّما وجد الفكرة أو الفرضية أنّها تكرّرت، كلّما اعتبرها شرعيّة وصحيحة. إذًا، هذا الذّكاء، أكثر من أنّه "توليديّ"، هو "داعم ومعزّز"، بمعنى أنّه يعيد ترتيب المحتويات الموجودة، ويساعد على ترسيخها، ولا يتحقّق غالبًا هل تحتوي على أخطاء أو أحكام مسبقة.

بهذه الطّريقة، ليس الخطر فقط في إعطاء شرعيّة لأخبار كاذبة، وتقوية الثّقافة المهَيمنة، بل الخطر أيضًا في أن نقوّض العملية التربويّة في مهدها. التّربية التي ينبغي أن تزوّد الطّلاب بإمكانيّة التّفكير الحقيقيّ، يمكن بذلك أن تصير تكرارًا للمفاهيم، التي سيتمّ تقييمها دائمًا أكثر على أنّها لا تقبل الجدل، وذلك ببساطة بسبب إعادة اقتراحها المستمرّ[11].

إعادة كرامة الإنسان إلى المركز من أجل رؤية أخلاقيّة مشتركة

يجب الآن أن نضيف إلى ما قيل سابقًا ملاحظة أكثر عموميّة. في الواقع، فإنّ وقت التّجديدات التّكنولوجيّة التي نشهدها اليوم يصاحبها وضع اجتماعيّ خاص وغير مسبوق: فقد أصبح من الصّعب بصورة متزايدة التّوصّل إلى اتفاقات بشأن القضايا الرّئيسيّة في الحياة الاجتماعية. وحتّى في الجماعات التي تتميّز ببعض الاستمراريّة الثّقافيّة، تنشأ أحيانًا مناقشات ومواجهات ساخنة تجعل من الصّعب إنتاج أفكار وحلول سياسيّة مشتركة تهدف إلى البحث عن ما هو جيّد وصحيح. وما عدا التّعقيد المشروع في الرّؤى التي تميّز الأسرة البشريّة، يظهر اليوم عامل يبدو أنّه مشترك بين هذه الحالات المختلفة، وهو غياب أو ضياع الحس الإنسانيّ وتضاؤل واضح لمفهوم الكرامة الإنسانيّة[12]. يبدو أنّ القيمة والمعنى العميق لإحدى الفئات الأساسيّة في الغرب قد ضاعت: وهي فئة الإنسان. ففي هذا العصر الذي تشكّك فيه برامج الذّكاء الاصطناعيّ في الإنسان وأعماله، فإنّ ضعف الرّوح المرتبطة بإدراك قيمة وكرامة الإنسان توشك بأن تكون أكبر نقطة ضعف في التّنفيذ والتّطوير لهذه الأنظمة. في الواقع، يجب ألاّ ننسى أنّه لا يوجد ابتكار محايد. لقد وُلدت التّكنولوجيا لهدف، وبتأثيرها على المجتمع البشريّ، تمثّل دائمًا شكلًا من أشكال النّظام في العلاقات الاجتماعيّة وموقعًا من السّلطة، وتجعل البعض قادرًا على القيام ببعض الأعمال وتمنع البعض الآخر من القيام بأعمال أخرى. هذا المكوّن السّلطوي في التّكنولوجيا يتضمّن دائمًا، بطريقة أكثر أو أقلّ وضوحًا، الرّؤية العالميّة للذين أنشأوها وطوّروها.

وهذا ينطبق أيضًا على برامج الذّكاء الاصطناعيّ. ولكي تكون هذه البرامج أدوات لبناء الخير ولبناء غد أفضل، لا بدّ من توجيهها دائما إلى خير كلّ إنسان. يجب أن يكون فيها إلهام أخلاقيّ.

القرار الأخلاقيّ، في الواقع، هو القرار الذي يأخذ بعين الاعتبار ليس فقط نتائج الفعل، بل أيضًا القِيَم المعرضة للخطر

ولهذا السّبب كان "نداء روما": في المصطلح "أخلاقيّات الخوارزميات" (algoretica) تكمن سلسلة من المبادئ تصلح أن تكون منصّة عالميّة ومتعدّدة قادرة على أن تجد الدّعم من الثّقافات والأديان والمنظمات الدّوليّة والشّركات الكبرى المدافعة عن هذا التّطوّر.

السّياسة التي نحتاج إليها

لذلك، لا يمكننا أن نخفي الخطر العمليّ، الكامن في الآليّة الأساسيّة التي بها يعمل الذّكاء الاصطناعيّ وهو أنّه يحصر رؤيّة العالم في واقع يمكن التّعبير عنه بالأرقام وبقوالب فكريّة معدّة مسبقًا، وبستثني مساهمة أشكال أخرى من الحقيقة والواقع، ويفرض نماذج أنثروبولوجيّة، اجتماعيّة واقتصاديّة، وثقافيّة موحدة. ومن ثمّ فإنّ النّموذج التّكنولوجيّ الذي يجسده الذّكاء الاصطناعيّ يوشك أن يفسح المجال لنموذج أكثر خطورة، والذي حدّدته من قبل باسم "النّموذج التّكنوقراطي"[15]. ولا يمكنّنا أن نسمح لأداة بمثل هذه القدرة، أي الذّكاء الاصطناعيّ، وفي الوقت نفسه لا نقدر أن نستغني عنها، بفرض مثل هذا النّموذج، بل يجب أن يكون الذّكاء الاصطناعيّ نفسه هو الأداة التي نستخدمها لمنع ذلك.

وهنا لا بدّ من العمل السّياسي، كما تذكّرنا الرّسالة العامّة، كلّنا إخوة- Fratelli tutti. بالتّأكيد "إنّ كلمة "سياسة" بالنّسبة للكثيرين اليوم هي كلمة مرفوضة، ولا يمكنّنا إلاّ أن نرى أنّ سبب ذلك هو وجود أخطاء لبعض السّياسيّين وفساد وعدم كفاءة. يُضاف إلى ذلك الاستراتيجيّات التي تسعى إلى إضعافها أو استبدالها بالاقتصاد أو الهيمنة عليها عبر بعض الأيديولوجيّات. ومع كلّ ذلك، هل يمكن للعالم أن يسير دون سياسة؟ هل يمكن أن نجد سبيلًا فعّالًا يؤدّي إلى الأخوّة الشّاملة والسّلام الاجتماعيّ دون سياسة صالحة؟"[16].

جوابنا على هذه الأسئلة الأخيرة هو: لا! السّياسة مطلوبة! أريد أن أؤكّد مجدّدًا في هذه المناسبة أنّه "أمام العديد من الأشكال السّياسية الضّيّقة أو السّاعية إلى الرّبح الفوري، [...] عظمة السّياسة تظهر، خصوصًا في الأوقات الصّعبة، حين يتمّ تطبيق المبادئ الكبيرة والتّفكير بالخير العام على المدى البعيد. لكنّ السّلطة السّياسيّة تجد صعوبة بالغة في قبول هذا الواجب ضمن مشروع وطني، وتحقيقه أصعب ضمن مشروع مشترك للبشريّة الحاليّة والمستقبليّة"[17].

السّيّدات الأعزّاء، السّادة الكِرام!

تفكيري هذا في تأثيرات الذّكاء الاصطناعيّ على مستقبل البشريّة يقودنا إلى النّظر في أهميّة "السّياسة السّليمة" حتّى ننظر إلى مستقبلنا بأمل وثقة. وكما قلت من قبل في مكان آخر، فإنّ "المجتمع العالميّ يعاني من أوجه قصور هيكليّة خطيرة لا يمكن حلّها بالتّرقيع أو بحلول سريعة عرضيّة. هناك أشياء يجب أن تتغيّر بعمليّة إعادة تفكير أساسيّة وتحوّلات رئيسيّة. السّياسة السّليمة وحدها تستطيع أن تقود هذا التّغيير، فتُشرك القطاعات المختلفة والمعرفة على تنوّعها. وبهذه الطّريقة، يستطيع الاقتصاد المندمج في مشروع سياسيّ واجتماعيّ وثقافيّ وشعبيّ، الذي يسعى إلى الخير العام، أن يفتح "الطّريق نحو فُرصٍ مُختلفةٍ، لا تستوجب الحدّ من الإبداع البشريّ ومن حلمه بالتّقدّم، بل تحتاج إلى توجيه هذه الطّاقة بأسلوب جديد" (*كُن مُسَبِّحًا*، 191)"[18].

وهذا الكلام ينطبق على الذّكاء الاصطناعيّ. ويجب على كلّ واحد أن يستخدمه بصورة صالحة، ويجب على السياسة أن تهيّئ الظّروف التي تجعل هذا الاستخدام الجيّد ممكنًا ومثمرًا.

شكرًا.

[1] *رسالة في اليوم العالمي السّابع والخمسين للسّلام*، الأوّل من كانون الثّاني/يناير 2024.

[2] راجع *المرجع نفسه*.

[3] راجع *المرجع السّابق*، 2.

[4] أشار البابا القدّيس بولس السّادس مُسبقًا إلى هذا التّناقض في كلمته إلى العامِلين في مركز:*Centro Automazione Analisi Linguistica dell'Aloysianum*، في 19 حزيران/يونيو 1964.

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

[6] رسالة عامّة بابويّة، *كُنْ مُسَبِّحًا* (24 أيّار/مايو 2015)، 102-114.

[7] *رسالة في اليوم العالمي السّابع والخمسين للسّلام*، الأوّل من كانون الثّاني/يناير 2024، 3.

[8] استنتاجات مارشال ماكلوهان وجون إم كولكين في ما يختص بنتائج استخدام الذّكاء الاصطناعيّ لها أهميّة واضحة.

[9] راجع *كلمة إلى المشاركين في الجلسة العامّة للأكاديميّة البابوية للحياة*، 28 شباط/فبراير 2020.

[10] راجع *رسالة في اليوم العالمي السّابع والخمسين للسّلام*، الأوّل من كانون الثّاني/يناير 2024، 4.

[11] راجع *المرجع نفسه*، 3 و 7.

[12] راجع دائرة عقيدة الإيمان، *إعلان كرامة الإنسان* (2 نيسان/أبريل 2024).

[13] راجع *كلمة إلى المشاركين في الجلسة العامّة للأكاديميّة البابوية للحياة*، 28 شباط/فبراير 2020.

[14] راجع *كلمة إلى المشاركين في مؤتمر "تعزيز كرامة الطّفل الرّقمية - من المفهوم إلى العمل"*، 14 تشرين الثاني/نوفمبر 2019؛ *كلمة إلى المشاركين في الجلسة العامّة للأكاديمية البابويّة للحياة*، 28 شباط/فبراير 2020.

[15] للحصول على تفسير أوسع، أشير إلى رسالتي العامّة "*كُنْ مُسَبِّحًا*" في العناية بالبيت المشترك الصّادرة في 24 أيار/مايو 2015.

[16] رسالة بابوية عامة، *كلّنا إخوة-Fratelli tutti*، في الأخوّة والصّداقة الاجتماعيّة (3 تشرين الأوّل/أكتوبر 2020)، 176.

[17] *المرجع السّابق*، 178.

[18] *المرجع السّابق*، 179.

Discorso pronunciato dal Santo Padre al G7

Grazie tante per la vostra attenzione.

Io qui ho due versioni: quella lunga e quella breve. Io leggerò la breve, soltanto.

L'intelligenza artificiale è uno strumento affascinante e tremendo

Saluto tutti Voi, Leader del Forum Intergovernativo del G7, saluto l'onorevole Giorgia Meloni, saluto la segretaria dell'Unione Europea e saluto presidenti, re, tutto quelli che sono presenti qui. La riflessione è "Gli effetti dell'intelligenza artificiale sul futuro dell'umanità".

«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)»[1]. La scienza e la tecnologia sono dunque prodotti straordinari del potenziale creativo di noi esseri umani[2].

Ebbene, è proprio dall'utilizzo di questo potenziale creativo che Dio ci ha donato che viene alla luce l'intelligenza artificiale.

Quest'ultima, come è noto, è uno strumento estremamente potente, impiegato in tantissime aree dell'agire umano: dalla medicina al mondo del lavoro, dalla cultura all'ambito della comunicazione, dall'educazione alla politica. Ed è ora lecito ipotizzare che il suo uso influenzerà sempre di più il nostro modo di vivere, le nostre relazioni sociali e nel futuro persino la maniera in cui concepiamo la nostra identità di essere umani[3].

Il tema dell'intelligenza artificiale è, tuttavia, spesso percepito come ambivalente: da un lato, entusiasma per le possibilità che offre, dall'altro genera timore per le conseguenze che lascia presagire.

Non possiamo, del resto, dubitare che l'avvento dell'intelligenza artificiale rappresenti una vera e propria rivoluzione cognitivo-industriale, che contribuirà alla creazione di un nuovo sistema sociale caratterizzato da complesse trasformazioni epocali. Ad esempio, l'intelligenza artificiale potrebbe permettere una democratizzazione dell'accesso al sapere, il progresso esponenziale della ricerca scientifica, la possibilità di delegare alle macchine i lavori usuranti; ma, al tempo stesso, essa potrebbe portare con sé una più grande ingiustizia fra nazioni avanzate e nazioni in via di sviluppo, fra ceti sociali dominanti e ceti sociali oppressi, mettendo così in pericolo la possibilità di una "cultura dell'incontro" a vantaggio di una "cultura dello scarto".

La portata di queste complesse trasformazioni è ovviamente legata al rapido sviluppo tecnologico dell'intelligenza artificiale stessa.

Proprio questo vigoroso avanzamento tecnologico rende l'intelligenza artificiale *uno strumento affascinante* e *tremendo* al tempo stesso ed impone una riflessione all'altezza della situazione.

In tale direzione forse si potrebbe partire dalla costatazione che l'intelligenza artificiale è innanzitutto *uno strumento*. E viene spontaneo affermare che i benefici o i danni che essa porterà dipenderanno dal suo impiego.

Questo è sicuramente vero, poiché così è stato per ogni utensile costruito dall'essere umano sin dalla notte dei tempi.

Questa nostra capacità di costruire utensili, in una quantità e complessità che non ha pari tra i viventi, fa parlare di una *condizione tecno-umana*: l'essere umano ha da sempre mantenuto una relazione con l'ambiente mediata dagli strumenti che via via produceva. Noi esseri umani, infatti, viviamo una condizione di ulteriorità rispetto al nostro essere biologico; siamo esseri sbilanciati verso il fuori-di-noi, anzi radicalmente aperti all'oltre. Da qui prende origine la nostra apertura agli altri e a Dio; da qui nasce il potenziale creativo della nostra intelligenza in

termini di cultura e di bellezza; da qui, da ultimo, si origina la nostra capacità tecnica. La tecnologia è così una traccia di questa nostra ulteriorità.

Tuttavia, l'uso dei nostri utensili non sempre è univocamente rivolto al bene. Anche se l'essere umano sente dentro di sé una vocazione all'oltre e alla conoscenza vissuta come strumento di bene al servizio dei fratelli e delle sorelle e della *casa comune* (cfr *Gaudium et spes*, 16), non sempre questo accade. Anzi, non di rado, proprio grazie alla sua radicale libertà, l'umanità ha pervertito i fini del suo essere trasformandosi in nemica di sé stessa e del pianeta[4]. Stessa sorte possono avere gli strumenti tecnologici. Solo se sarà garantita la loro vocazione al servizio dell'umano, gli strumenti tecnologici riveleranno non solo la grandezza e la dignità unica dell'essere umano, ma anche il mandato che quest'ultimo ha ricevuto di "coltivare e custodire" (cfr *Gen* 2,15) il pianeta e tutti i suoi abitanti. Parlare di tecnologia è parlare di cosa significhi essere umani e quindi di quella nostra unica condizione tra libertà e responsabilità, cioè vuol dire parlare di etica. Non si può scartare una cosa dall'altra.

E questo è vero in misura maggiore per l'intelligenza artificiale, la quale è uno strumento ancora più complesso. Direi quasi che si tratta di uno strumento *sui generis*. Conviene sempre ricordare che la macchina può, in alcune forme e con questi nuovi mezzi, produrre delle scelte algoritmiche. Ciò che la macchina fa è una scelta tecnica tra più possibilità e si basa o su criteri ben definiti o su inferenze statistiche. L'essere umano, invece, non solo sceglie, ma in cuor suo è capace di decidere. Per questa ragione, di fronte ai prodigi delle macchine, che sembrano saper scegliere in maniera indipendente, dobbiamo aver ben chiaro che all'essere umano deve sempre rimanere la decisione, anche con i toni drammatici e urgenti con cui a volte questa si presenta nella nostra vita. Condanneremmo l'umanità a un futuro senza speranza, se sottraessimo alle persone la capacità di decidere su loro stesse e sulla loro vita condannandole a dipendere dalle scelte delle macchine. Abbiamo bisogno di garantire e tutelare uno spazio di controllo significativo dell'essere umano sul processo di scelta dei programmi di intelligenza artificiale: ne va della stessa dignità umana.

Proprio su questo tema permettetemi di insistere: in un dramma come quello dei conflitti armati è urgente ripensare lo sviluppo e l'utilizzo di dispositivi come le cosiddette "armi letali autonome" per bandirne l'uso, cominciando già da un impegno fattivo e concreto per introdurre un sempre maggiore e significativo controllo umano. Nessuna macchina dovrebbe mai scegliere se togliere la vita ad un essere umano.

Rimettere al centro la dignità della persona in vista di una proposta etica condivisa

A quanto già detto va ora aggiunta un'osservazione più generale. La stagione di innovazione tecnologica che stiamo attraversando, infatti, si accompagna a una particolare e inedita congiuntura sociale. Si registra come uno smarrimento o quantomeno un'eclissi del senso dell'umano e un'apparente insignificanza del concetto di dignità umana[5]. Ed è così che in questa stagione in cui i programmi di intelligenza artificiale interrogano l'essere umano e il suo agire, proprio la debolezza dell'*ethos* connesso alla percezione del valore e della dignità della persona umana rischia di essere il più grande *vulnus* nell'implementazione e nello sviluppo di questi sistemi. Non dobbiamo dimenticare infatti che nessuna innovazione è neutrale: nessuna innovazione è neutrale. La tecnologia nasce per uno scopo e, nel suo impatto con la società umana, rappresenta sempre una forma di ordine nelle relazioni sociali e una disposizione di potere, che abilita qualcuno a compiere azioni e impedisce ad altri di compierne altre. Questa costitutiva dimensione di potere della tecnologia include sempre, in una maniera più o meno esplicita, la visione del mondo di chi l'ha realizzata e sviluppata.

Questo vale anche per i programmi di intelligenza artificiale. Affinché questi ultimi siano strumenti per la costruzione del bene e di un domani migliore, debbono essere sempre ordinati al bene di ogni essere umano. Devono avere un'ispirazione etica.

La decisione etica, infatti, è quella che tiene conto non solo degli esiti di un'azione, ma anche dei valori in gioco e dei doveri che da questi valori derivano. Per questo ho salutato con favore la firma a Roma, nel 2020, della *Rome Call for AI Ethics*[6] e il suo sostegno a quella forma di moderazione etica degli algoritmi e dei programmi di intelligenza artificiale che ho chiamato "algoretica" [7]. In un contesto plurale e globale, in cui si mostrano anche sensibilità diverse e gerarchie plurali nelle scale dei valori, sembrerebbe difficile trovare un'unica

gerarchia di valori. Ma nell'analisi etica possiamo ricorrere anche ad altri tipi di strumenti: se facciamo fatica a definire un solo insieme di valori globali, possiamo però trovare dei principi condivisi con cui affrontare e sciogliere eventuali dilemmi o conflitti del vivere.

La politica di cui c'è bisogno – perché c'è bisogno di una politica su questo.

Non possiamo nascondere il rischio concreto, poiché insito nel suo meccanismo fondamentale, che l'intelligenza artificiale limiti la visione del mondo a realtà esprimibili in numeri e racchiuse in categorie preconfezionate, estromettendo il rapporto e l'apporto di altre forme di verità e imponendo modelli antropologici, socio-economici e culturali uniformi. Il paradigma tecnologico incarnato dall'intelligenza artificiale rischia allora di fare spazio a un paradigma ben più pericoloso, che ho già identificato con il nome di "paradigma tecnocratico [si sbaglia, dice democratico]"[8].

Ed è proprio qui che è urgente l'azione vostra, l'azione politica, come ricorda l'Enciclica *Fratelli tutti*. Certamente «per molti la politica oggi è una brutta parola, e non si può ignorare che dietro questo fatto ci sono spesso gli errori, la corruzione, l'inefficienza di alcuni politici. Non di tutti: alcuni. A ciò si aggiungono le strategie che mirano a indebolirla, a sostituirla con l'economia o a dominarla con qualche ideologia. E tuttavia, può funzionare il mondo senza politica? No. Può trovare una via efficace verso la fraternità universale e la pace sociale senza una buona politica?»[9].

La nostra risposta a queste ultime domande è: no! La politica serve! Mi viene in mente quello che un Papa ha detto sulla politica: "È la forma più alta della carità, è la forma più alta dell'amore". La politica serve. Voglio ribadire in questa occasione che «davanti a tante forme di politica meschine e tese all'interesse immediato […] la grandezza politica si mostra quando, in momenti difficili, si opera sulla base di grandi principi e pensando al bene comune a lungo termine. Il potere politico fa molta fatica ad accogliere questo dovere in un progetto di Nazione e ancora di più in un progetto comune per l'umanità presente e futura»[10]. Sempre c'è la tentazione di uniformare tutto. Mi viene in mente un romanzo, famoso, dell'inizio del 1900, "The Lord of the World", un romando inglese futurista, che fa vedere il futuro senza politica, il futuro soltanto uniformante. È bello leggerlo, è interessante.

Gentili Signore, Illustri Signori!

Questa mia riflessione sugli effetti dell'intelligenza artificiale sul futuro dell'umanità ci conduce così alla considerazione dell'importanza della "sana politica" per guardare con speranza e fiducia al nostro avvenire. Come ho già detto altrove, «la società mondiale ha gravi carenze strutturali che non si risolvono con rattoppi o soluzioni veloci meramente occasionali. Dobbiamo andare alle radici. Ci sono cose che devono essere cambiate con reimpostazioni di fondo e trasformazioni importanti. Solo una sana politica potrebbe averne la guida, coinvolgendo i più diversi settori e i più vari saperi. In tal modo, un'economia integrata in un progetto politico, sociale, culturale e popolare che tenda al bene comune può "aprire la strada a opportunità differenti, che non implicano di fermare la creatività umana**,** né l'intelligenza artificiale e il suo sogno di progresso, ma piuttosto di incanalare tale energia in modo nuovo" (*Laudato si'*, 191)» [11].

Questo è proprio il caso dell'intelligenza artificiale. Spetta ad ognuno farne buon uso e spetta alla politica creare le condizioni perché un tale buon uso sia possibile e fruttuoso.

Grazie.

---

[1] *Messaggio per la LVII Giornata Mondiale della Pace del 1° gennaio 2024*, 1.

[2] Cfr *ibid*.

[3] Cfr *ivi*, 2.

[4] Lett. enc *Laudato si'* (24 maggio 2015), 102-114.

[5] Cfr Dicastero per la Dottrina della Fede, Dichiarazione *Dignitas infinita* circa la dignità umana (2 aprile 2024).

[6] Cfr *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[7] Cfr *Discorso ai partecipanti al Convegno "Promoting Digital Child Dignity – From Concet to Action"*, 14 novembre 2019; *Discorso ai partecipanti alla Plenaria della Pontificia Accademia per la Vita*, 28 febbraio 2020.

[8] Per una più ampia esposizione, rimando alla mia Lettera Enciclica *Laudato si'* sulla cura della casa comune del 24 maggio 2015.

[9] Lettera enc. *Fratelli tutti* sulla fraternità e l'amicizia sociale (3 ottobre 2020), 176.

[10] *Ivi*, 178.

[11] *Ivi*, 179.

[01033-IT.01] [Testo originale: Italiano – Trascrizione di lavoro]

[B0504-XX.03]