



La Santa Sede

PAPA FRANCESCO PARTECIPA ALLA SESSIONE DEL G7 SULL'INTELLIGENZA ARTIFICIALE
[13-15 giugno 2024]

DISCORSO DEL SANTO PADRE FRANCESCO

*Borgo Egnazia (Puglia)
Venerdì, 14 giugno 2024*

[[Multimedia](#)]

- **[Versione integrale del Discorso preparato dal Santo Padre](#)**
- **[Discorso pronunciato dal Santo Padre al G7](#)**

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, entusiasmo 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 constatazione 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.

Sofisticato 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 Concept 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.