

Digeat N.3 - 19 Settembre 2024

# La fiducia come chiave per la costruzione di un'IA antropocentrica

Di Alexandra Lisac



## Abstract

La fede nella scienza è cosa diversa dalla fiducia nella tecnologia. Il discorso del Santo Padre al G7 fa da filo conduttore a questo articolo, che propone una breve esplorazione della differenza concettuale che sussiste tra questi due termini in rapporto all'attuale era digitale, con uno sguardo di apertura sull'AI Act ed alla convenzione-quadro del Consiglio d'Europa recentemente firmata dalla Commissione Europea e da alcuni Paesi extra-UE.

## Indice

- Il discorso del Papa al G7
- Fede e fiducia nella tecnologia
- Dall'AI ACT alla convenzione-quadro sull'IA
- Conclusioni

Il Santo Padre pone da tempo grande attenzione ai temi legati alle nuove tecnologie, in particolare quelle governate dall'IA, ugualmente al centro del discorso tenuto in occasione della sua partecipazione al G7.<sup>[1]</sup>

## Il discorso del Papa al G7

La riflessione del Papa si incentra proprio **sull'impatto dell'IA sul futuro dell'umanità** ed evidenzia come l'avvento di questa potente e nuova tecnologia rappresenti **una rivoluzione di tipo cognitivo-industriale**, tale da contribuire alla creazione di un nuovo sistema sociale caratterizzato da trasformazioni epocali.<sup>[2]</sup>

Il Santo Padre ha posto l'accento su come le scelte algoritmiche a base statistica sulle quali si fonda l'IA **sfuggano al controllo degli stessi sviluppatori** ed è, pertanto, necessario garantire uno spazio di controllo significativo all'essere umano.

Papa Francesco ci ricorda con grande chiarezza che l'uomo deve sempre avere potere decisionale e che non dobbiamo mai dimenticare che l'essere umano è sempre in evoluzione ed è capace di sorprendere con le sue azioni, cosa che la macchina non potrà mai essere in grado di fare.

Per il Santo Padre, sarebbe un grave errore non rammentare che l'IA altro non è che una macchina, non può proporre principi generali e non è né oggettiva, né neutrale, né immune da errori e pregiudizi.

## Fede e fiducia nella tecnologia

Le parole del Santo Padre ci ricordano che **parlare di tecnologia vuol dire parlare di etica**, in quanto attiene all'essere umano ed al potenziale creativo della sua intelligenza. Nel dibattito sulle innovazioni tecnologiche di IA, ci si imbatte sovente in due termini, **fede e fiducia**, che esprimono due concetti diversi, da non confondere tra loro.

La fede attiene ad una dimensione intima, rimanda ad un rapporto con la tecnologia quasi sacrale. La fiducia, invece, si riferisce ad una dimensione concreta, poiché significa credere in qualcosa o in qualcuno.

Se guardiamo alla sua etimologia, la parola fiducia deriva dal **latino fides (pistis in greco)**. La fides e la pistis per il mondo classico **rappresentavano delle vere e proprie divinità**, che avevano il compito di regolare i rapporti fra i cittadini e le leggi dello Stato.

È nel Medioevo che questi due termini **si intrecciano con il foedus, il patto**, che, sigillato dal giuramento, andava a regolamentare non soltanto i rapporti politico-sociali, ma anche quelli personali tra individui.<sup>[3]</sup>

Ricondurre il tema dell'intelligenza artificiale al concetto di fiducia significa ancorarlo alla concretezza e conseguentemente alla riflessione su come costruire un ambiente digitale ed un ecosistema tecnologico etico e incentrato sul rispetto dell'essere umano e della sua dignità.

Se si pensa all'**IA generativa con riferimento agli LLM (Large Language Model)** come il celeberrimo ChatGPT, è inevitabile riferirsi al prompt, ossia all'input che viene fornito ai modelli da parte dell'utente.

Il prompt influisce in grande misura sull'output finale, pertanto, più questo è ben definito, chiaro e preciso, più attendibile e affidabile sarà la risposta finale, ovviamente sempre nei limiti di eventuali *bias* da black box. L'attività di prompting, **si basa sulle domande e da esse si sviluppa a seconda di come queste domande vengono formulate e strutturate**.

I software di intelligenza artificiale rispondono sempre e comunque alle domande, perché **non possono non fornire soluzioni**, in quanto è proprio per generare risposte che essi sono stati progettati.

L'IA degli LLM è ontologicamente un solutore dei più disparati quesiti a cui non può rispondere con un generico "non lo so", poiché, come ci ricorda il Santo Padre, ***l'IA è un utensile che, essenzialmente, è disegnato per la risoluzione di un problema e funziona per mezzo di un concatenamento logico di operazioni algebriche su dati raffrontati per individuare le correlazioni.***<sup>[4]</sup>

Ciò considerato, si comprende come il tema della fiducia sia di cruciale importanza nell'ambito dell'intelligenza artificiale, quindi, secondo una sorta di metaforica attività di prompting, bisognerebbe domandarsi non cosa vogliamo dall'IA, ma come vogliamo che questa tecnologia sia e rimanga in futuro.

Ragionare in questi termini significa dare alla fiducia un ruolo centrale nelle politiche di governo dell'intelligenza artificiale, come ribadito dalle istituzioni UE nel processo di elaborazione dell'AI Act e nella recente convenzione-quadro.

## Dall'AI ACT alla convenzione-quadro sull'IA

I principi alla base dell'AI Act trovano il loro fondamento nei pareri dei Comitati Nazionali di Bioetica.<sup>[5]</sup>

In particolare, i pareri resi dal CNB/CNBB invitano ad evitare eccessive speranze ed eccessivi timori e suggeriscono di adottare un atteggiamento di fiducia e di cautela, promuovendo l'etica nel disegno delle tecnologie (*ethics by design/in design/for designers*), assicurando una tecnologia che sia orientata ad incorporare i valori e assicurare la centralità del paziente.<sup>[6]</sup>

L'AI Act, adottando **un approccio proporzionato basato sul rischio**, ha posto come obiettivo la promozione di un'intelligenza artificiale affidabile e coerente con i valori e gli interessi dell'Unione Europea. Recentemente, la Commissione Europea ha licenziato la **convenzione-quadro del Consiglio d'Europa sull'IA, i diritti umani e lo Stato di diritto, sottoscritta anche da Paesi extra-UE**, con l'obiettivo di provvedere a costruire un approccio comune per un'IA compatibile con i valori democratici, in linea con l'AI Act ed in coerenza con gli obblighi di protezione dei diritti umani nel rispetto del diritto internazionale. Il tema della fiducia assume significativa rilevanza centrale, in quanto la convenzione-quadro prevede adeguati requisiti di qualità e di sicurezza per tutta la durata del ciclo di vita dei sistemi di IA, la supervisione della sperimentazione e dello sviluppo affidata alle autorità competenti, nonché l'obbligo di adottare misure di mitigazione dei rischi e di valutazioni sugli impatti potenziali e reali sui diritti umani, sulla democrazia e sullo Stato di diritto.<sup>[7]</sup>

## Conclusioni

L'UE dimostra, così, di rimanere saldamente ancorata all'adozione di una **prospettiva antropocentrica** che trova il suo perno nel concetto di fiducia, che, per poter essere concretamente ed efficacemente costruita deve, necessariamente, prevedere sistemi di IA che pongano al centro il rispetto della dignità dell'essere umano.

---

## NOTE

[1] Papa Francesco, *Discorso al G7, 14.6.2024; si rinvia inoltre alla lettura delle encicliche Laudato Si, Laudate Deum e del Messaggio per la 57esima Giornata Mondiale per la Pace.*

[2] Ibidem

[3] Ibidem

[4] Cfr. Discorso al G7 del Pontefice.

[5] Tra tutti si ricordano: il Nuffield Council on Bioethics, *Artificial Intelligence in Healthcare and Research 2020*; lo Swedish National Council on Medical Ethics, *Artificial Intelligence in Healthcare, 2020*; il Comitato Nazionale per la Bioetica (CNB), il Comitato per le biotecnologie biosicurezza e scienze della vita (CNBB), *Intelligenza Artificiale e medicina: aspetti etici, 2020.*

[6] Parere del CNB/CNBB, *Intelligenza Artificiale e medicina: aspetti etici, 2020.*

[7] Convenzione-quadro sull'intelligenza artificiale, i diritti umani e lo Stato di diritto del Consiglio d'Europa del 17.5.2024.