Il monito di Yoshua Bengio, tra i padri dell’intelligenza artificiale: «Prendereste un aereo che ha il 10% di probabilità di cadere?»

«Quando ho iniziato a studiare l’intelligenza artificiale nel 1985 ero davvero affascinato. Non immaginavo i progressi degli ultimi decenni e la rapidità con cui si sarebbe evoluta. Nemmeno prevedevo il tipo di impatto che avrebbe avuto sul mondo. E lasciate che vi dica: stiamo vedendo solo la punta dell’iceberg. Se continua cosi potrebbe andare molto meglio. Ma anche molto peggio…».

Inizia così il discorso di **Yoshua Bengio,** uno dei padri dell’intelligenza artificiale, informatico di adozione

canadese, arrivato a Roma nei giorni scorsi per partecipare al tavolo degli esperti mondiali dell'Intelligenza artificiali, organizzato dal giornalista **Riccardo Luna.** L'occasione è il ***World Meeting on human Fraternity***, promosso dalla Fondazione Fratelli tutti, con la Basilica di San Pietro e il dicastero per lo Sviluppo umano integrale, per rispondere a una domanda essenziale: "Come essere umani nell'era dell'intelligenza artificiale?".

Sessantun'anni, Bengio è considerato **uno dei "padri fondatori" del deep learning**, ha sviluppato metodi che hanno insegnato alle macchine ad apprendere. È autore di un elenco infinito di cose belle. È uno degli scienziati più citati al mondo, quello con il più alto numero di citazioni scientifiche nel campo dell'intelligenza artificiale.

Yoshua Bengio, foto credit @No Panic

Nato a Parigi, ha **vissuto da sempre in Canada**. Laurea in Ingegneria elettrica, un master e dottorato di ricerca in Informatica, studia reti neurali e riconoscimento vocale. Diventa professore all'**Université de Montréal** nel 1993. Poi

inizia a vincere premi su premi. Nel 2018 conquista il Premio Turing, considerato il **Nobel dell'informatica,** con Geoffrey Hinton e Yann LeCun. Nel 2025 la rivista *Time* lo inserisce nella lista delle "100 persone più influenti nel campo dell'AI". Sempre quest'anno ritira il **Queen Elizabeth Prize for Engineering**, uno dei riconoscimenti più prestigiosi a livello internazionale nel campo dell'ingegneria.

Insomma, un mito. Che a un certo punto nella sua vita decide di guardare dentro quello che aveva contributo a creare. «Quando ho visto ChatGPT tre anni mi sono posto molte domande. E ho deciso di **concentrarmi totalmente sui rischi e le minacce dell'Intelligenza artificiale.** Lo faccio per i miei figli, i miei nipoti. Per i vostri figli. L'errore più grande che le persone fanno è immaginare il futuro dell'Intelligenza artificiale solo come una piccola estensione di ciò che stiamo vedendo ora». Non è cosi. Non sarà così.

«Stiamo costruendo macchine che ci sorpasseranno in molti campi. Pensiamo agli agenti AI, capaci di decidere in autonomia. Hanno una conoscenza super avanzata e questo è grandioso. Vedremo sistemi che ci aiutano a risolvere molti problemi. Ma… la teoria ci sta mostrando che se hanno un obiettivo non allineato ai nostri, potrebbero decidere di perseguirlo con ostinazione, qualunque siano le conseguenze per noi.

E la triste verità è che la scienza, le big tech, le università **non sanno come costruire sistemi che siano allineati a noi e non danneggino gli esseri umani**. Sistemi che possono decidere di ostacolarci, ingannarci e mentire per preservare se stessi. Andando contro le nostre istruzioni. Non è fantascienza, svegliatevi!»

A questo punto, nella sala delle Scuderie di Palazzo Altieri a Roma dove si tiene il *The artificial intelligence Table*, cala il silenzio. Anzi, il silenzio diventa ancora più pesante. E Bengio continua: «Stiamo costruendo macchine che un giorno

potrebbero competere con noi ed essere più intelligenti di noi. Lo vogliamo davvero?».

Sono seduta tra un giornalista di Radio Rai, Massimo Cerofolini, e l'head of Euroasiatica news, Carlo Marino. Davanti a noi c'è tutta la comunità dei giornalisti tech. L'evento è a porte chiuse e solo su invito. Ci guardiamo, come a chiederci lo vogliamo davvero?

Bengio sembra capirlo e prova a rassicurarci. «L'AI può produrre benefici enormi ma solo se la si guida saggiamente. Per questo ho deciso di dedicare il resto della mia carriera a questo problema nella speranza di imparare alcune cose».

Bengio ha fondato, infatti, **LawZero,** una non profit che si dedica allo sviluppo di un' AI sicura e sotto controllo umano. Il nome richiama Legge Zero del libro Isaac Asimov. Che dice pressappoco cosi. Un robot non può recare danno all'umanità o permettere che l'umanità subisca danno. Finora, **LawZero ha raccolto 30 milioni di dollari**. Tra i donatori: **Jaan Tallinn** (co-fondatore di Skype), **Eric Schmidt** (ex CEO di

Google), **Open Philanthropy**, il Future of Life Institute e la Gates foundation. LawZero sta sviluppando una "Scientist AI", un sistema di AI progettato per dare priorità all'onestà.

«Sono uno scienziato e un ricercatore del Michigan. Ho bisogno di trovare una soluzione tecnica a queste due domande. La prima: come **progettiamo l'IA senza danneggiare le persone?** La seconda è: **come governiamo quel potere, se lo costruiamo?** In questo periodo sto lavorando a un progetto globale che coinvolge Cina, USA, e Unione Europea e devo ammettere che gli scienziati non sono sempre d'accordo. Ma se non riusciamo a collaborare, **l'AI può essere usata come strumento di dominio.** Da chi vuole più potere, da chi vuole generare caos, dai terroristi o da persone con ideologie strane».

Come affrontare tutto questo?

«L'unico modo è gestirla come bene pubblico globale». Qualcuno vicino a me, in sala, dice sottovoce: illusioni. «Sì, non è ciò che sta accadendo» – continua – «Stiamo vedendo una corsa, una folle competizione tra i vari Paesi e le varie aziende, dove sicurezza e etica non vengo preservate. Per affrontare questi rischi, ci vuole una leadership forte, morale». Qui Bengio fa riferimento ai leader religiosi che possono essere cruciali in questo momento. Poi continua:

«Dobbiamo creare un AI che serve all'umanità non un'umanità al servizio dell'AI. La posta in gioco è alta. **Continueranno a esistere l'umanità, le democrazie, la pace?** Controlleremo ancora il nostro futuro?».

Silenzio.

«Anche se ci fosse solo l'1% di possibilità che uno qualsiasi di questi rischi si materializzi, dovremmo essere estremamente cauti. Inoltre un gran numero di ricercatori pensa che la probabilità di tali rischi sia **molto più alta dell'1%**».

Poi si ferma, si rivolge a tutti noi presenti e ci chiede:

«Salireste su un aereo che ha il 10% di probabilità di cadere? Probabilmente no. Ma la cattiva notizia è che solo poche persone al mondo decideranno per noi se spingerci oltre e prendere quell' aereo…»

Applausi. Bengio scende dal palco, entra in video collegamento da Toronto Geoffrey Hinton, premio Nobel per la Fisica, in qualche modo maestro di Bengio e il **primo tra i tecnici a lasciare al mondo l'allarme**. Ma questa è un'altra storia che vi racconterò presto.