

Next Generation Startup

RECORD DI CONTRIBUTI

Investitori privati e fondi credono nel progetto

Clearbox AI è la startup tech italiana che aiuta le aziende a lanciare progetti di intelligenza

artificiale e di Analytics, generando dati sintetici di alta qualità. La startup ha ottenuto 720 mila euro di contributi, da investitori privati e fondi europei, impiega oggi sette persone: tre soci lavoratori, tre dipendenti, un collaboratore.

Clearbox è stata recentemente selezionata dalla Commissione Europea per il progetto Women TechEU, che supporta le startup deep-tech guidate da donne per valorizzare il talento e favorire l'innovazione nell'ecosistema tech europeo.

La proprietà intellettuale è riconducibile alla fonte specificata in testa alla pagina. Il ritaglio stampa è da intendersi per uso privato



Clearbox AI aiuta le aziende a sviluppare analisi più precise

Studiando i Big Data si ottengono servizi e prodotti migliori

di **Letizia Magnani**

Dal metaverso i dati "sintetici" per proteggerci dalle frodi

E SE PERSINO I DATI venissero creati con il metaverso? È questa la domanda che si sono posti alcuni giovani, fra i quali spicca **Shalini Kurapati (nella foto sopra)**, Co-Founder e Ceo di **Clearbox AI**. Una start up che si occupa proprio di questo. Ovvero di Data Engine. I dati sintetici sono creati a immagine di quelli reali, e possono già da ora, favorire performance più accurate dei modelli di intelligenza artificiale. Non si tratta di un futuro lontano, ma immediato. Tanto che si stima che già dal prossimo anno oltre il 60% dei dati utilizzati in progetti di intelligenza artificiale saranno generati sinteticamente. La tecnologia corre veloce, se si pensa che nel 2021 solo l'1% dei dati era non reale. Questa tecnologia può essere applicata a tutti i campi che necessitano di molti dati per migliorare i processi: assicurazioni, finanza, energia, telecomunicazioni, mobilità urbana, retail. Società come Amazon, American Express, John Deere li stanno già utilizzando, per essere più precisi e per aggirare i termini della privacy, soprattutto in Europa.

Da dove nasce l'idea di Clearbox AI?

«Nel 2019 da un'idea mia e di altri tre co-founder. Ho conosciuto alcuni di loro in Olanda dove, dopo la triennale in ingegneria in India, ho vinto una borsa di studio del governo olandese per un master a Delft dove ho svolto anche il dottorato. È lì che ho iniziato la mia attività imprenditoriale, con una prima società di consulenza sull'intelligenza artificiale. Ora vivo a Torino».

Dall'Olanda all'Italia con una idea innovativa?

«In Olanda ho avuto l'occasione di notare che c'erano ancora molti problemi con questa tecnologia, specialmente per quanto riguarda la reperibilità e gestione dei dati. È facile creare modelli, ma spesso è difficile renderli robusti e calibrati perché i dati con cui sono allenati risultano essere sporchi e sbilanciati e così succede che rimangono in laboratorio o non funzionano. Da qui l'idea di creare una soluzione per sfruttare la potenza dell'intelligenza artificiale in modo responsabile. Così quando abbiamo immaginato **Clearbox AI** ab-

biamo deciso di lasciare tutto ed è stato naturale scegliere Torino. All'epoca collaboravo già con il Politecnico e quando abbiamo presentato l'idea di impresa all'incubatore I3P ci hanno subito risposto e poi accettato».

Cosa fate concretamente?

«Aiutiamo le aziende a lanciare progetti di intelligenza artificiale e di Analytics, attraverso la generazione di dati sintetici di alta qualità. La missione aziendale consiste nel comprendere e risolvere le sfide che le imprese incontrano nello sviluppo dei processi di Intelligenza Artificiale. Questi ostacoli sono spesso legati ai dati sensibili che sono difficili da gestire a livello di privacy e che non sono abbastanza rappresentativi per tutte le fasce di popolazione, o la loro quantità non è sufficiente per garantire risultati di successo. Per tutti questi motivi, i nostri dati sintetici possono essere utili a diverse professionalità nelle aziende: reparti di data science, innovazione, software engineering e privacy».

Cosa sono i dati sintetici?

«Se guardiamo al mondo dell'intelligenza artificiale e alla sua progressiva adozione, solo 1 progetto su 10 riesce ad arrivare alla fase di produzione. Questo accade per molte ragioni che passano dall'organizzazione, alle aspettative sui progetti, ma soprattutto dai dati. Problemi legati ai dati significa mancanza di accesso alle informazioni decisive, perché magari si tratta di dati sensibili, perché sono sporchi, sono parziali o sono alterati dai pregiudizi di chi li ha sviluppati. Il dato sintetico è una soluzione nata per risolvere questo tipo di problemi. Sono dati artificiali, generati da algorit-

60%

La percentuale dei dati utilizzati in progetti di intelligenza artificiale saranno generati sinteticamente dal 2024

ALGORITMI CON UN SEME DI REALTÀ

Solo un progetto su dieci realizzati con l'IA si trasforma in realtà. L'uso dei dati sintetici è fondamentale per arrivare a risultati statisticamente attendibili. Qui a fianco, il team della startup: Luca Cavaletto, Alessandra Nicolosi, Carmine d'Amico, Andrea Minieri, Luca Gilli, Shalini Kurapati, Matteo Giovannetti



mi, però con un seme di realtà. Non vengono dal nulla, ma ricreano in modo matematico situazioni della vita vera partendo da dati reali, mantenendo le stesse proprietà statistiche e predittive di quelli originali. Quindi risultano realistici».

Come si producono i dati?

GLI ASPETTI POSITIVI

«Oltre al rispetto della privacy questa tecnologia permette di avere più elementi di analisi quando questi sono limitati e corregge i dataset sbilanciati»

«Nascono dal nostro Data Engine fondato su tecnologia proprietaria e agnostica basata su modelli generativi avanzati, creata anche grazie alle solide radici del team nel mondo della ricerca. La soluzione supporta le grandi aziende sia ad incrementare la disponibilità di dati, la loro qualità e quindi mitigando eventuali bias interni ai dati, ma anche per aiutarle in termini di policy, compliance e privacy degli stessi».

Quali sono i principali benefici dei dati sintetici?

«Sono tre: permettono di tutelare la privacy e allo stesso tempo sviluppare modelli, forniscono più dati quando questi sono quantitativamente limitati mantenendone l'utilità, e correggono i dataset che sono sbilanciati a favore di un gruppo. Questi benefici possono venire utilizzati in tanti ambiti. Abbiamo tante richieste dai settori bancari, dalle assicurazioni, dalla finanza e anche dal settore energetico e dal marketing. Nel settore bancario, per esempio, lavoriamo nel rilevamento di frodi. In questo campo, c'è una forte componente di sbilanciamento dei dati: le banche non hanno abbastanza casi di frodi accertate per riuscire ad allenare modelli di Intelligenza Artificiale per il loro rilevamento automatico. Con i dati sintetici possono allenare i modelli su un dataset più ricco, ottenendo dei miglioramenti delle performance degli stessi anche del 15%-20%».

La tecnologia cambia velocemente, cosa significa per una startup come la vostra?

«Continueremo a lavorare per accelerare l'adozione dell'intelligenza artificiale responsabile attraverso un approccio data-centric, cioè che mette al centro il dato, e anche per creare e valutare dataset che siano performanti, che preservino la privacy, ma anche equi e rappresentativi di tutti i gruppi della società».

© RIPRODUZIONE RISERVATA

