

DATI SINTETICI

Clearbox AI, “lavatrice” di dati

Rendere robusti e calibrati i modelli di IA è una questione complessa, a causa dei dati “sporchi” con cui sono addestrati. La startup torinese aiuta le imprese generando dati sintetici di alta qualità

DIGIULIA CIMANELLI

In Italia soltanto un'impresa su cinque è guidata da una donna. Secondo dati di Unioncamere, a fine settembre 2022 le aziende femminili erano soltanto il 22,18% delle imprese italiane. Se si scende nel dettaglio delle startup innovative del settore It, la percentuale di imprenditrici sul totale è ancora più bassa. Le imprese guidate da persone nate fuori dall'Italia poi, sono ancora meno: circa il 10% del totale, e tra le startup innovative la percentuale scende drasticamente.

Anche per questo quella di **Clearbox AI** è una storia d'impresa che rompe gli schemi. A fondare

una società che si occupa di dati sintetici e quindi opera con l'intelligenza artificiale è stata, in Italia, una giovane donna indiana, Shalini Kurapati. Dopo aver completato gli studi in ingegneria in India ha ottenuto una borsa di studio del governo olandese per un master a Delft, dove ha anche conseguito un dottorato. Lì ha scoperto la sua passione imprenditoriale, fondando una società di consulenza sull'intelligenza artificiale.

Durante il suo percorso, Shalini Kurapati ha notato un problema diffuso nel campo dell'IA.

«La difficoltà nel reperire e gestire dati di alta qualità. Creare modelli di intelligenza artificiale è diventato facile, ma renderli robusti e cali-

brati è risultato un compito arduo a causa dei dati “sporchi” e sbilanciati con cui venivano addestrati. Così mi

è venuta l'idea di creare una soluzione per sfruttare l'intelligenza artificiale in modo responsabile».

Decisa a trasformare il suo sogno in realtà, Kurapati e i suoi co-fondatori hanno preso una decisione coraggiosa: lasciare tutto e trasferirsi a Torino, dove collaboravano già con il Politecnico.

Cosa fa esattamente Clearbox AI? «La nostra azienda - prosegue la fondatrice - aiuta le imprese a lanciare progetti di intelligenza artificiale e analisi dati generando dati sintetici di alta qualità».

I dati sintetici sono dati artificiali, generati da algoritmi, ma con un seme di realtà. Non vengono creati dal nulla, ma «ricreano situazioni reali in modo matematico, mantenendo le stesse proprietà statistiche e predittive dei dati originali» continua Kurapati.

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, i cosiddetti “bias”.

«Noi abbiamo tante richieste dai settori bancari, dalle assicurazioni, dalla finanza e anche dal settore energetico e dal marketing - spiega ancora la cofondatrice di **Clearbox AI** -. Nel settore bancario, per esempio, lavoriamo nel ri-

levamento 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 automati-

co. Con i dati sintetici possono allenare i modelli su un dataset più ricco, ottenendo miglioramenti delle performance anche del 15-20%».

Ma non è tutto, conclude Shalini Kurapati: «Siamo anche in grado di risolvere problemi legati ai bias nei confronti delle donne e delle minoranze. Prendiamo l'esempio di Amazon: l'algoritmo dell'azienda discriminava il curriculum delle donne rispetto a quelli degli uomini. Questo, si è scoperto, succedeva perché l'algoritmo prendeva in considerazione dati storici nei quali le donne non erano rappresentate all'interno del mondo del lavoro. Il problema, insomma, si nascondeva in un dataset che doveva essere aggiornato. Con il nostro modello questo non sarebbe accaduto». —

© RIPRODUZIONE RISERVATA





Shalini Kurapati, indiana, ha co-fondato **Clearbox AI**