Intelligenza artificiale, ora Dall-E 2 sa anche creare le proteine: i possibili sviluppi

Usare e riutilizzare ciò che già c'è per ottenere qualcosa di più evoluto è una delle filosofie cardine di ogni branca dell'ingegneria. Lasciando che questa forma mentis entri nei laboratori della farmacopea si ottengono risultati interessanti. In questo caso, la tecnologia su cui si erige **Dall-E 2** è usata alcune da aziende e atenei americani per realizzare proteine diverse da quelle esistenti.

Non desta stupore che Dall-E 2, sistema di Intelligenza artificiale (AI) creato nei laboratori della statunitense OpenAI e capace di realizzare immagini partendo da un testo, si stia prestando a diversi settori, tra i quali la moda e la cinematografia, fornendo immagini strane e meravigliose on request.

È molto più sensazionale che la stessa tecnologia alla base di Dall-E 2 sia stata recentemente utilizzata in alcuni laboratori di biotecnologia, che hanno iniziato a utilizzare questo tipo di AI generativa per creare **nuove tipologie** di proteine non presenti in natura.

I primi annunci

Due aziende del settore hanno recentemente annunciato, in maniera quasi parallela e separata, l'avvio di **programmi** che utilizzano modelli di diffusione per generare progetti di nuove proteine con una precisione mai vista prima. Generate Biomedicines, una startup a stelle e strisce con sede a Somerville, Massachusetts, ha presentato un programma chiamato

Chroma che l'azienda afferma essere il **Dall-E 2 della biologia**.

Allo stesso tempo, un team statunitense della University of Washington ha realizzato un programma parimenti simile a Dall-E 2: RoseTTAFold Diffusion. In quest'ultimo caso è stato dimostrato che il modello è stato in grado di generare progetti per nuove proteine che possono poi essere **realizzate in laboratorio**, tutte diverse da quelle esistenti.

Questi generatori artificiali di proteine possono essere indirizzati a **produrre protidi** con proprietà specifiche, con forme, dimensioni e funzioni diverse. In effetti, è possibile creare nuove proteine per un determinato fine, come può essere lo sviluppo di nuovi farmaci più efficaci di quelli esistenti.

Le proteine sono i "mattoni" fondamentali che costituiscono gli esseri viventi. Negli animali (e, quindi, nell'uomo) sono deputate, per esempio, a digerire il cibo, a **contrarre i muscoli** e a rilevare la luce. Le proteine sono i "bersagli" principali dei farmaci, quando ci si ammala svolgono un ruolo importante nel sistema immunitario e, molti dei farmaci più recenti, sono proprio a base di proteine. Attualmente **chi progetta farmaci** deve attingere a un elenco di ingredienti composto da proteine naturali. L'obiettivo della generazione di proteine mediante AI è quello di estendere questo elenco con un complesso **quasi infinito** di proteine progettate artificialmente.

L'innovazione

Va da sé che le tecniche alla base della progettazione di queste proteine **non sono nuove**. Tuttavia, gli approcci precedenti sono stati più lenti e non sono stati in grado di progettare proteine di grandi dimensioni o complessi proteici (macchine molecolari composte da più proteine accoppiate tra loro). E queste proteine sono spesso cruciali per il trattamento di determinate malattie.

Chroma e RoseTTAFold Diffusion, peraltro, non sono nemmeno il primo esempio di utilizzo di modelli di diffusione per la generazione di proteine. Entrambe sono, però, classificabili come i **primi programmi completi** in grado di produrre progetti precisi per un'ampia varietà di proteine.

I modelli di diffusione di proteine sono reti neurali (che riflettono quelle umane) addestrate a rimuovere le perturbazioni casuali aggiunte loro da diversi input. Dato un **insieme casuale** di pixel, un modello di diffusione cerca di trasformarlo in un'immagine riconoscibile.

In Chroma, il rumore viene aggiunto disfacendo le catene di aminoacidi di cui è composta una proteina. Dato un gruppo casuale di queste catene, Chroma cerca di metterle insieme per formare, appunto, una proteina. Guidato da vincoli specifici sull'aspetto del risultato, Chroma può **generare nuove proteine** con proprietà specifiche.

Con RoseTTAFold Diffusion, invece, la partita si gioca diversamente. Il team della University of Washington adotta un **approccio diverso** ma con risultati finali simili. Il proprio modello di diffusione parte da una struttura ancora più confusa. Un'altra differenza fondamentale è che RoseTTAFold Diffusion utilizza le informazioni su come i pezzi di una proteina si incastrano tra loro, fornite da una rete neurale separata addestrata a prevedere la struttura delle proteine (come fa AlphaFold di DeepMind). E tutto ciò guida il processo generativo complessivo.

I risultati ottenuti

Peraltro, sia Chroma che RoseTTAFold Diffusion mostrano una serie impressionante di risultati. Entrambi i sistemi sono in grado di generare proteine con diversi gradi di simmetria, tra cui proteine circolari, triangolari o esagonali.

Per illustrare la versatilità del loro programma, Chroma ha

generato proteine a forma delle ventisei lettere dell'alfabeto (latino) e a forma dei numeri da zero a dieci. Entrambi i team, però, possono anche generare pezzi di proteina con un'unica forma, anche abbinando nuove parti di proteine a strutture esistenti. Certo è che la maggior parte di queste nuove strutture artificiali, non avrebbe alcuna utilità nella pratica (siamo ovviamente ancora agli albori). Ma, come si è detto, poiché la **funzione di una proteina** è determinata dalla sua forma, è fondamentale poter generare strutture diverse on request.

Ovviamente, generare strani progetti a macchina è una cosa magari lo si fa per testare a fondo il sistema. Altra cosa è trasformare questi progetti in "proteine reali" e, quindi, in farmaci.

Le verifiche

Per verificare se Chroma possa produrre progetti effettivamente realizzabili, Generate Biomedicines ha preso sequenze di alcuni dei suoi progetti (stringhe di aminoacidi che compongono la proteina) e le ha sottoposte a un altro programma di Intelligenza artificiale. Il team di Somerville ha scoperto che il 55% di tali sequenze ripiega nella struttura generata da Chroma, il che suggerisce che si tratta di progetti per **proteine potenzialmente realizzabili**.

Il team della University of Washington ha eseguito un test simile, anche se – stando a ciò che lo stesso ha dichiarato – si è spinto molto più in là di Generate Biomedicines nella valutazione del proprio modello. Per il team della University of Washington, il risultato principale è la generazione di una nuova proteina che si attacca all'ormone paratiroideo, che controlla i livelli di calcio nel sangue. Stando sempre a quanto dichiarato, a RoseTTAFold Diffusion è stato fornito un ormone e nient'altro, con la richiesta di creare una proteina che si legasse ad esso.

Quando a Seattle (dove ha sede la University of Washington) hanno testato la nuova proteina in laboratorio, hanno scoperto che la stessa si legava all'ormone in modo più stretto di qualsiasi altra cosa che avrebbe potuto essere generata con altri metodi computazionali.

Da Generate Biomedicines si riconosce che l'invenzione di nuove proteine è solo il primo passo di molti altri. In fin dei conti la stessa azienda opera nel settore farmaceutico e ciò che conta realmente è riuscire a produrre farmaci che funzionino. I farmaci a base di proteine devono essere prodotti in grandi quantità, poi testati in laboratorio e infine somministrati a **esseri umani**. Tutto questo può richiedere anni.

Se si accelerasse questo processo, anche in una sola di queste fasi, il mondo intero ne beneficerebbe nel medio-lungo termine. Vedremo l'evolversi della materia, nei prossimi mesi, a cosa ci porterà.