

Tecno - *Collasso dell'Intelligenza Artificiale: l'antidoto è la "sorpresa"*

Pisa - 13 nov 2026 (Prima Notizia 24) **Studio Cnr-Isti: dare priorità ai testi meno prevedibili durante l'addestramento di un modello linguistico aiuta a limitare il model collapse, il fenomeno per cui i testi generati diventano via via meno vari e meno coerenti durante l'autofagia.**

La diffusione capillare dei modelli di IA generativa aumenta sempre di più il rischio dell'autofagia, il processo in cui modelli di IA vengono addestrati su output prodotti da loro stessi. Questo circolo vizioso produce una degenerazione nota come collasso del modello (model collapse), nel quale si degrada la qualità del contenuto generato. Uno studio curato dall'Istituto di scienza e tecnologie dell'informazione del Consiglio Nazionale delle Ricerche di Pisa (CNR-ISTI), Università di Pisa, Scuola Normale Superiore e Victoria University di Wellington (Nuova Zelanda) analizza il problema proponendo una strategia per costruire sistemi di IA generativa più solidi. La ricerca, pubblicata sulla rivista ACM Transactions on Intelligent Systems and Technology, individua nella "sorpresa" l'elemento chiave di addestramento per preservare la diversità dei testi e la capacità creativa. "Analizzando quattro modelli di IA generativa - come Llama, sviluppato da Meta - abbiamo osservato che il collasso può manifestarsi in modi diversi. Da un lato, il contenuto generato diventa sempre meno vario e presenta sempre le stesse parole; dall'altro, le frasi generate sono meno coerenti e significative e le risposte meno sensate", spiega Daniele Gambetta, ricercatore post-doc presso l'Università di Pisa. "La soluzione comune è distinguere i testi umani da quelli prodotti dall'IA, selezionando solo quelli umani per l'addestramento. Tuttavia, è sempre più difficile fare questa distinzione man mano che i modelli diventano più sofisticati. Il nostro metodo propone invece di misurare la perplexity, cioè il livello di sorpresa nei documenti secondo il modello. Un documento è sorprendente per il modello se il modello stesso l'avrebbe scritto in modo diverso", aggiunge Luca Pappalardo, ricercatore del CNR-ISTI e co-direttore dello studio: "I testi meno sorprendenti contribuiscono di più al collasso se usati nell'addestramento del modello. Nei nostri esperimenti abbiamo dimostrato che selezionando per l'addestramento solo i documenti sorprendenti si mitiga il problema del collasso". "Nell'autofagia il modello smette di apprendere nuove informazioni. Reintrodurre la sorpresa nell'addestramento richiama il modo in cui i bambini imparano dagli eventi inattesi", prosegue il professor Dino Pedreschi. "Questo lavoro si inserisce nel contesto più ampio della coevoluzione umano-IA, un ambito di ricerca che studia come i sistemi artificiali si trasformano e vengono trasformati dall'ecosistema informativo in cui operano". "Questo approccio non richiede di sapere se un documento è stato scritto da un umano o da una macchina e, in un'epoca in cui distinguere i due tipi di contenuto è sempre più difficile, questa proprietà è fondamentale per la strategia", conclude Gizem Gezici, ricercatrice presso la Scuola Normale Superiore. "Il nostro lavoro fornisce una possibile soluzione a un problema che diventerà sempre più impellente con la diffusione di questi modelli nella produzione testuale in vari contesti".

(Prima Notizia 24) Venerdì 13 Novembre 2026

Sede legale : Via Costantino Morin, 45 00195 Roma
E-mail: redazione@primanotizia24.it