

CHATBOT E IDEOLOGIA

Ernie: l'algoritmo "comunista" e il futuro che (non) vogliamo per l'IA

Home > Cultura E Società Digitali



Gli algoritmi non sono "di destra" o "di sinistra", discriminatori o estremisti. Sono solo istruzioni sequenziali di calcolo che elaborano dati secondo determinate regole. La loro apparente "ideologia" dipende quindi interamente dalla programmazione e uso da parte degli esseri umani. Perciò, il caso del chatbot cinese Ernie solleva dubbi sulla censura e l'ideologia incorporata nell'AI

Pubblicato il 27 nov 2023

Stefano Pietropaoli

Università di Firenze – Officina informatica CRID, Unimore – SC Centro Studi

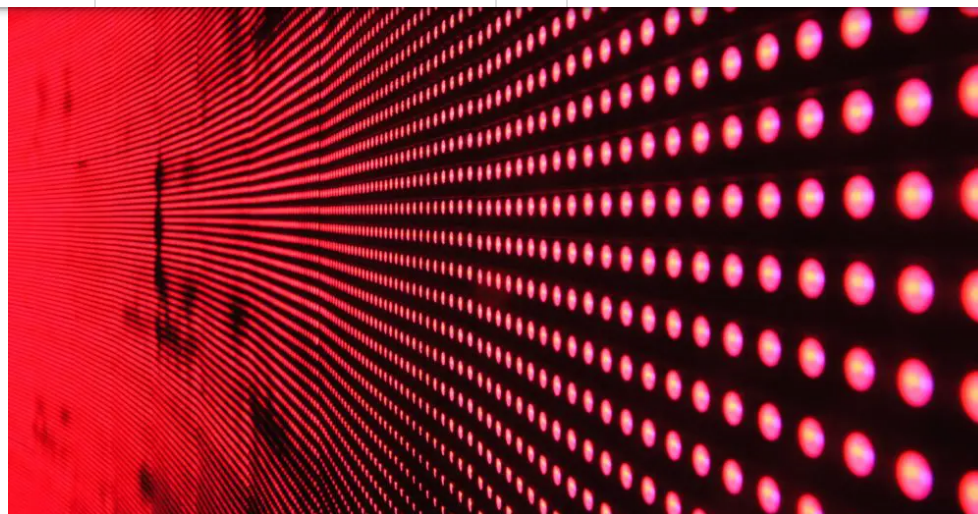
Andrea Simoncini

SC Centro Studi



Quella che indichiamo con l'espressione "**intelligenza artificiale**" – tanto seducente quanto pericolosa nella sua apparente semplicità – è una costellazione estremamente complessa di tecnologie, anche diversissime tra loro, che perseguono obiettivi pratici diversi (come tradurre un testo, guidare un veicolo oppure formulare una diagnosi medica) ma restano accomunate da un unico scopo.

Scopo che altro non è che **la realizzazione di sistemi artificiali capaci di eseguire compiti che "tipicamente" richiedono l'intelligenza di esseri umani**, simulando alcune delle capacità umane di percezione, apprendimento, ragionamento, decisione e risoluzione dei problemi.



Ecco il chatbot cinese, nuova arma di controllo sociale

23 Marzo 2023

di Riccardo Berti e Franco Zumerle

Indice degli argomenti

- IA generativa, LLM e chatbot
- Ernie: un chatbot comunista?
- La disciplina cinese dei servizi di IA generativa
- Un'IA conforme ai valori del socialismo
- Perché dobbiamo prendere sul serio i chatbot cinesi
- È sempre "garbage in, garbage out"
- Governo dei dati: come si sta muovendo l'Europa
- La forza di persuasione individuale e collettiva dell'IA

IA generativa, LLM e chatbot

Tra le più stupefacenti espressioni del campo di ricerca sull'IA (e, più precisamente, nell'ambito della "IA generativa") troviamo oggi i sistemi basati su **modelli di lingua su larga scala (LLM)**, capaci di simulare una conversazione con un essere umano (chatbot, o meglio *Conversational Generative Artificial Intelligence Chatbot*).

job creeper: i consigli dell'Osservatorio HR Innovation del Poli

 Recruiting  Lavoro agile

[Leggi l'informativa sulla privacy](#)

E-mail aziendale*

Acconsento alla comunicazione dei miei dati a [terzi](#) affinché li trattino per proprie finalità di marketing tramite modalità automatizzate e tradizionali di contatto.

SCARICA ORA

Il **GPT** (Generative Pre-trained Transformer), sviluppato da OpenAI e alla base del funzionamento di ChatGPT, è sicuramente il più noto di questi modelli linguistici, ma non l'unico. Praticamente tutti i big player dell'intelligenza artificiale hanno già realizzato o stanno elaborando modelli simili. Basti pensare a Google, coinvolta direttamente nello sviluppo di **BERT** (*Bidirectional Encoder Representations from Transformers*), **XLNet** (sviluppato da Google insieme alla Carnegie Mellon University), T5 (*Text-to-Text Transfer Transformer*), fino ad arrivare a **PaLM** (*Pathways Language Model*), su cui si basa il funzionamento di Bard, oggi forse il principale antagonista di ChatGPT.

Ernie: un chatbot comunista?

Tutti i modelli citati finora "parlano" inglese e sono stati sviluppati direttamente o indirettamente negli Stati Uniti d'America (e, come risulta evidente sin dal nome, anche CamemBERT, addestrato in francese, altro non è che una variante di uno dei modelli sviluppati da Google). Esiste tuttavia un'eccezione (destinata ad essere presto affiancata da altre), rappresentata da **Ernie** (o meglio: Ernie 3.0-Titan, *Enhanced Representation through Knowledge Integration*), il modello linguistico sviluppato dal colosso cinese Baidu su cui si basa l'omonimo chatbot, rilasciato nel marzo 2023 e pianamente operativo dal 31 agosto, quando ha raggiunto 1 milione di utenti a sole 19 ore dal lancio.

la pena ricordarlo, ChatGPT di OpenAI non può essere né essere scaricata da utenti cinesi né essere "agganciata" a software sviluppati in Cina).

Prima di approdare sul mercato, Baidu ha dovuto ottenere l'**autorizzazione dal governo** di Pechino. Una rapida ricerca su Wikipedia tocca subito il nodo della questione: «Ernie Bot è soggetto al regime di censura del governo cinese». Nelle note si rinvia ad alcuni articoli che sottolineano l'**elusività di Ernie di fronte a domande "politicamente sensibili"**. Elusività che diventa in alcuni casi palese reticenza. «Possiamo parlare di tutto quello che volete», dice all'inizio. Per poi proseguire: «Ma vi preghiamo di notare che alcuni argomenti possono essere delicati o toccare questioni legali e sono quindi soggetti alla vostra responsabilità».

Posto davanti alla domanda su un tema delicato – come il genocidio degli **uiguri** nello Xinjiang o le **proteste** di Hong Kong del 2019-2020 – il sistema risponde: «Parliamo di qualcos'altro»; oppure: «Cambiamo argomento e ricominciamo». Nel caso di **Piazza Tienanmen**, Ernie dice di non disporre di informazioni rilevanti. Ma non si tratta soltanto di non rispondere e di provare a cambiare argomento. Si tratta anche di prendere posizioni estremamente "nette". Basti fare l'esempio di Taiwan: per Ernie «fa parte del territorio sacro della Repubblica Popolare Cinese. La sovranità e l'integrità territoriale della Cina non possono essere violate o divise».

Siamo, dunque, di fronte a un chatbot comunista?

La risposta potrebbe essere positiva, se la domanda fosse posta correttamente. Ma **le cose sono più complicate di quanto potrebbero sembrare** a un primo sguardo. Baidu, insieme ad altre aziende specializzate nell'IA generativa (SenseTime, Baichuan Intelligent Technology, Zhipu AI), ha dovuto sottoporre al governo cinese un *risk assesment* per il proprio prodotto e dimostrare **la conformità alle linee guida formulate dall'Amministrazione cinese per il cyberspazio**. Proprio queste disposizioni meritano un approfondimento.

La disciplina cinese dei servizi di IA generativa

Il Provvedimento Provvisorio sulla Gestione dei Servizi di Intelligenza Artificiale Generativa, emanato con decreto n. 15 del 23 maggio 2023 dall'Ufficio per l'Amministrazione delle Informazioni su Internet (con l'approvazione del Comitato Nazionale per lo Sviluppo e le Riforme e di vari

l'interesse pubblico», protegga i diritti dei cittadini, delle persone giuridiche e di altre organizzazioni. Si tratta, ovviamente, di una disposizione suscettibile di un'interpretazione molto ampia, ma che non ci sorprenderebbe ritrovare in documenti analoghi formulati da istituzioni europee o statunitensi.

Allo stesso modo, molti dei principî generali elencati all'**articolo 4** del Provvedimento potrebbero ben figurare nelle linee guida dei paesi occidentali: l'IA generativa non deve essere impiegata per promuovere il terrorismo, l'estremismo, l'odio razziale, la discriminazione, la violenza, la pornografia e le fake news.

Più precisamente, il Provvedimento richiede l'adozione, durante il processo di progettazione dell'algoritmo e di selezione dei dati di addestramento, di **misure in grado di evitare discriminazioni basate su razza, religione, nazionalità, regione, genere, età, occupazione e salute**.

Inoltre, vengono sottolineati la centralità della tutela della proprietà intellettuale e il divieto di impiegare algoritmi, dati e piattaforme per attuare pratiche monopolistiche o di concorrenza sleale. Insieme alla tutela della reputazione, della onorabilità, della riservatezza dei dati personali, anche la protezione della salute, fisica e mentale, è richiamata in più punti, e in particolare all'articolo 9, che stabilisce l'obbligo per i fornitori di definire con chiarezza obiettivi, contesto e uso dei servizi offerti, permettendo agli utenti di utilizzare l'generativa in modo "scientifico e razionale" e adottando **misure efficaci per prevenire la dipendenza degli utenti**, e in particolare dei minori, dai servizi di intelligenza artificiale generativa.

Un altro richiamo importante è quello alla "**trasparenza**" dei servizi di intelligenza artificiale generativa, presentata come indispensabile per migliorare l'accuratezza e l'affidabilità dei contenuti generati. Durante la prestazione dei servizi, i fornitori devono espressamente "etichettare" i contenuti generati, come immagini e video, in conformità a una articolata serie di prescrizioni tecniche.

Per quanto riguarda i **dati di addestramento**, l'articolo 7 stabilisce che devono essere adottate tutte le misure possibili per migliorarne la qualità, la veridicità, l'accuratezza, l'obiettività e la varietà. Possono essere utilizzati soltanto **dati acquisiti in maniera legale**, nel rispetto dei diritti di proprietà intellettuale, e in caso di dati personali occorre ottenere il consenso degli interessati o un'altra base giuridica prevista dalla legge.

statunitense.

Un'IA conforme ai valori del socialismo

Tuttavia, ce n'è una che, invece, difficilmente potremmo trovare in altri paesi. Oltre ai principi che abbiamo richiamato, **l'articolo 3 stabilisce che le applicazioni basate sull'IA generativa devono «aderire ai valori fondamentali del socialismo»** e non possono generare contenuti «che incitino alla sovversione del potere statale e all'abbattimento del sistema socialista».

L'impostazione di fondo, dunque, è chiara e dichiarata: **l'IA generativa deve sviluppare applicazioni «positive e salutari», cioè conformi ai valori socialisti.** In questa prospettiva si comprende meglio il senso degli articoli 5 e 6, che invitano gruppi industriali, aziende, istituti di istruzione e ricerca e istituzioni culturali a “collaborare” nella ricerca in materia di IA generativa, nello sviluppo di algoritmi, framework, chip e piattaforme software correlate, e nella gestione dei dati di addestramento, anche tramite la costruzione condivisa di infrastrutture di base e di piattaforme pubbliche di dati di addestramento: «Deve essere promosso il coordinamento per migliorare l'efficienza dell'utilizzo delle risorse di calcolo. Deve essere promossa l'apertura graduale e organizzata di dati pubblici di addestramento di alta qualità. Deve essere promossa l'adozione di dati sicuri e affidabili».

Sulla “sicurezza” di questi strumenti è però *l'articolo 17* che dispone la norma fondamentale. I servizi di intelligenza artificiale generativa che possono influenzare l'opinione pubblica o che possono produrre effetti di mobilitazione sociale devono essere sottoposti a **una valutazione preventiva di sicurezza**: devono essere registrati, modificati o cancellati in conformità con le norme sulla gestione dei servizi informativi su Internet. È, questa, una declinazione particolarissima del concetto di cybersecurity su cui vale la pena riflettere.

Ai sensi dell'articolo 19, le autorità competenti hanno poi il diritto di “ispezionare” i servizi di intelligenza artificiale generativa, mentre i fornitori devono cooperare, fornendo spiegazioni sulla fonte, la scala, il tipo, le regole di etichettatura, i meccanismi algoritmici e i dati di addestramento.

Chi viola queste disposizioni va incontro alle **sanzioni previste all'articolo 21**. La misura minima è qualcosa che potremmo chiamare “ammonimento”, con cui si prescrive di apportare determinate

anche di natura penale.

Perché dobbiamo prendere sul serio i chatbot cinesi

Sarebbe sin troppo facile liquidare la normativa cinese sui servizi di IA generativa come l'anacronistico, protervo e sfacciato tentativo di imporre nell'era digitale modi e forme del sovietismo novecentesco. **Occorre invece prendere sul serio il caso dei chatbot cinesi per due ragioni: la prima è che il caso di Ernie mette in evidenza i pericoli che queste tecnologie rappresentano per lo Stato di diritto, la democrazia, i diritti umani (e non solo per i valori socialisti!); la seconda è che, guardando all'esperienza cinese senza pregiudizi, possiamo riflettere sul modo occidentale di sviluppare l'IA generativa:** un'occasione per capire se siamo sulla strada giusta, e se siamo davvero consapevoli di ciò che stiamo facendo.

Può far sorridere che proprio la Cina denunci i rischi di manipolazione dell'opinione pubblica insiti in strumenti come ChatGPT ed Ernie Bot. Ma **la preoccupazione che i chatbot possano giocare un ruolo geopoliticamente rilevante è più che fondata.** Guardandoci nello specchio del comunismo algoritmico, dunque, dobbiamo prendere sul serio prima di tutto la questione della **rilevanza dei dati di addestramento** degli strumenti di IA generativa. La soluzione cinese della limitazione e del filtraggio dei dati che alimentano gli algoritmi presenta molti svantaggi, non solo sul piano della tutela di diritti e libertà fondamentali, ma anche su quello squisitamente tecnologico (meno dato significa inibizione dello sviluppo). Tuttavia, dobbiamo prendere atto che tocca un nodo delicatissimo che non può essere lasciato irrisolto.

È sempre "garbage in, garbage out"

Il caso cinese, infatti, pone in evidenza uno dei fattori più complessi per chi si occupa di regolazione dell'Intelligenza Artificiale: per garantire la protezione dei diritti o di valori giuridici come la non discriminazione o la trasparenza, con riferimento agli "output" di questi sistemi tecnologici, occorre agire sull' "input". È il vecchio slogan degli esperti di programmazione, espresso dall'acronimo GIGO (garbage in, garbage out): **se inserisci spazzatura, esce spazzatura.**

Le valutazioni, i testi, le predizioni, **le decisioni che prenderà il software dipendono dalla qualità dei dati su cui è addestrato** – oltre che da come esso viene addestrato (ovvero dai fattori di controllo

Governo dei dati: come si sta muovendo l'Europa

Per questo motivo, il progetto di regolamentazione europea dell'Intelligenza Artificiale (il progetto di *AI Act*), con un passo in avanti di grande rilievo in materia di regolazione, dedica un articolo espressamente al tema: l'art. 10, intitolato, "**Dati e governo dei dati**".

Questo articolo prevede che i sistemi di Intelligenza Artificiale ad alto rischio (dunque, quelli oggetto di regolazione) e "che si avvalgono di tecniche che comportano l'addestramento di modelli sviluppati sulla base di insiemi di dati di addestramento, convalida e prova", dovranno soddisfare specifici requisiti di qualità.

In particolare, dovranno indicare **con quali criteri e secondo quali scelte sono gestiti i dati di addestramento**; come sono stati collezionati, annotati, classificati, "puliti", "arricchiti", aggregati; dovranno, altresì, consentire l'esame in vista dei possibili "biases".

Ma tra questi requisiti il più incisivo ed anche il più discusso, è quello previsto dal **comma 3: "I set di dati di formazione, convalida e test devono essere pertinenti, rappresentativi, privi di errori e completi"**.

L'Europa, dunque, analogamente alla Cina, sposta con forza l'attenzione sui dati di addestramento.

Lo scopo, però, è opposto: non solo non impone una visione ufficiale o un orientamento – l'adesione ai valori fondamentali del socialismo – ma richiede che i dati utilizzati siano il più possibile rappresentativi, privi di errori e completi, potremmo tradurre impropriamente, "**imparziali**".

Ovviamente, così facendo si addossa un peso notevole all'industria che produce ovvero impiega sistemi di AI ad alto rischio. Da un lato, molti specialisti ritengono estremamente difficile e costoso riuscire ad avere dataset effettivamente "completi" ovvero, "privi di errori", ritenendo i bias inevitabili; dall'altro, trova nuova linfa quello specifico settore della data-science, chiamata, "data cleansing", specializzata nella "pulizia" dei data-set dai dati errati o incompleti.

La forza di persuasione individuale e collettiva dell'IA

L'altro tema che questa vicenda fa emergere – "**come in uno specchio**"* – è quello della forza di persuasione individuale e collettiva, potremmo dire, che possiedono i sistemi di intelligenza

servizi tecnologici che producono informazioni, descrizioni, immagini, musica e più in genere forme di espressione culturale. Non a caso lo stesso articolo prosegue affermando **che le norme sulla IA si applicheranno solo se non in contrasto con la disciplina già esistente in Cina in materia di editoria, produzione cinematografica e televisiva, creazione letteraria e artistica.**

Torna alla mente la notissima affermazione leniniana ripresa, non a caso da Mussolini, **"il cinema è l'arma più forte"**⁴, per ricordare quanto gli strumenti di cultura di massa, proprio per la loro capacità educativa, possano diventare potentissime armi di "propaganda" nei regimi totalitari.

Ragion per cui il decreto del governo cinese, si propone di agire su due versanti, uno negativo ed uno positivo.

Sul primo, esso stabilisce che l'IA generativa non deve essere impiegata per "promuovere il terrorismo, l'estremismo, promuovere l'odio etnico, la discriminazione etnica, la violenza, la pornografia, nonché informazioni false e dannose e altri contenuti vietati dalle leggi e dai regolamenti amministrativi"; finalità che, come dicevamo, potremmo ritrovare abbastanza agevolmente anche in atti legislativi europei (ma che già solleverebbero forti perplessità nel sistema costituzionale americano – notoriamente ispirato alla massima protezione del First Amendment -)

Ma quel che preoccupa maggiormente è il versante "positivo" del divieto, quello, cioè in cui il Governo pone un obbligo di contenuto: la IA generativa deve "aderire ai valori fondamentali del socialismo e non incitare alla sovversione del potere statale o a rovesciare il sistema socialista"

Qui, come si vede, l'obbligo non è solo ad evitare certi tipi di contenuti o finalità, ma si pone **una finalità attiva**, istituendo un vero e proprio controllo di "conformità ideologica" delle risposte fornite da questi sistemi.

Orbene, è facile immaginare che, quantomeno in Europa, questa seconda prospettiva di strumentalizzazione propagandistica delle piattaforme di IA generativa trovi **una ferma quanto unanime opposizione**; ciononostante, la questione cinese sollecita una riflessione ulteriore: se in Cina l'obiettivo dichiarato è quello di difendere i valori socialisti, dobbiamo pure prendere atto che anche i chatbot occidentali possono egualmente essere "indirizzati", seppur in una direzione ideologicamente diversissima. Basti ricordare che OpenAI è stata accusata da commentatori

questo o quell'acquisto. Potremmo dire, dunque, che questi algoritmi se non aderiscono "in pieno ai valori del socialismo", certamente sposano quelli del capitalismo.

Crediamo, a questo riguardo, sia importante sottolineare che **gli algoritmi, di per sé, non sono "di destra" o "di sinistra", discriminatori ovvero estremisti:** non sono razzisti o xenofobi, comunisti o capitalisti. Non hanno orientamenti politici o ideologici intrinseci. Sono semplicemente istruzioni sequenziali di calcolo che elaborano dati secondo determinate regole o procedure.

Un algoritmo può essere progettato per implementare principi ed obiettivi politici specifici, compresi quelli associati a determinate ideologie (ad esempio, potrebbe essere sviluppato un algoritmo che calcola e distribuisce risorse seguendo i principi comunisti di condivisione delle risorse). **Ma la loro apparente "ideologia" dipende interamente dalla decisione, progettazione, programmazione e utilizzazione da parte degli esseri umani.** Il mezzo (anche se meglio sarebbe impiegare, con Bernard Stiegler, la categoria platonica del *pharmakon*) non è in sé buono o cattivo. È il fine per cui viene impiegato che ci consente di prendere posizione sulla sua bontà o malvagità.



Per questo, una delle possibili ma non auspicabili derive dello sviluppo delle tecnologie di IA generativa potrebbe essere il pullulare di chatbot rispondenti ciascuno a differenti convinzioni politiche, orientamenti culturali, preferenze sessuali, credenze.

A ciascuno il suo (chatbot), verrebbe da dire, parafrasando Sciascia.

* Ingmar Bergman (o San Paolo ... per speculum et in aenigmate)

 **WHITEPAPER**

Talent retention: 5 punti su cui lavorare (ed errori da evitare) per avere dipendenti soddisfatti e

 Data Management  Data warehousing

[Leggi l'informativa sulla privacy](#)

19 Dic 2023

19 Dic 2023

Agenda Digitale

Registro dati sanitari: come costruirlo in linea col GDPR



Scaricalo gratis! DOWNLOAD


Scaricalo gratis!

DOWNLOAD

Argomenti


F Fake News

F formazione

 Tutto su Cyber Security

T Tutto su GDPR

Canali

 Cultura e società digitali



LA SURVEY

Pnrr e digitale accoppiata vincente per il 70% delle pmi italiane

Articoli correlati




IL FRONTE ANTITRUST

"Google monopolio illecito", tutti i punti del grande processo USA

15 Set 2023

di Barbara Calderini

Condividi 




INTELLIGENZA ARTIFICIALE

Sogei: "Ecco la nostra strategia per diventare un'azienda cognitiva"

04 Lug 2023

di **Monica Gabrielli**


Condividi 

LE RICERCHE

Far crescere le prestazioni e il fatturato? Respirare con consapevolezza aria pulita

06 Giu 2023

di Francesco Russo

Condividi 



Scaricalo gratis!

DOWNLOAD