TECNOLOGIA

I rischi e le opportunità della nuova IA condivisa

VINCENZO AMBRIOLA


AGORÀ

TECNOLOGIA

I rischi e le opportunità della nuova IA condivisa

Il diritto internazionale prova a dare regole ma l'etica rimane centrale. Il G7: costruire ambientiinformatici aperti, partendo dai dati da usare per l'addestramento iniziale delle macchine Il 15marzo, a Trento, si è conclusa la riunione ministeriale "Industria, tecnologia e digitale", sotto lapresidenza italiana del G7, con l'approvazione di una dichiarazione sulle regole di governodell'intelligenza artificiale. I membri del G7 hanno espresso la volontà di sviluppare i sistemi diintelligenza artificiale in modo etico e coerente con i principi che sono alla base della democrazia,per favorire la coesione, la resilienza e il benessere della società. La dichiarazione include unasezione dedicata all'intelligenza artificiale nel settore pubblico, in cui si auspica la costruzionedi un ambiente "aperto" e abilitante per lo sviluppo e il rilascio di sistemi sicuri e affidabili. Inlinea con lo spirito del "Regolamento" recentemente approvato dal Parlamento Europeo, la dichiarazioneribadisce l'approccio basato sull'analisi dei rischi dei sistemi di intelligenza artificiali.

Ma cosa significa, esattamente, realizzare un ambiente aperto? La dichiarazione non entra neldettaglio di questa affermazione ma possiamo provare a darne un'interpretazione plausibile, basandocisu concetti ampiamente condivisi dalla comunità informatica. Partiamo dal fatto che un sistema diintelligenza artificiale è realizzato mediante programmi software e successivamente addestratoutilizzando una gran quantità di dati, con il coinvolgimento di molte persone che ne verificano e necorreggono il funzionamento. Ad esempio, per insegnare a un sistema a riconoscere i gatti è necessarioraccogliere e usare moltissime immagini e poi assicurarsi che funzioni correttamente. Il sistema nonsi deve far confondere dall'immagine di un cane, durante il processo di riconoscimento di un gatto. Iprogrammi software usati possono essere disponibili pubblicamente, e in questo caso si parla disoftware aperto, oppure di proprietà esclusiva di chi li ha scritti, e in questo caso si parla disoftware chiuso o proprietario. Anche i dati di addestramento possono essere pubblici o proprietari.

L'auspicio del G7 di costruire un ambiente aperto e abilitante implica, quindi, la disponibilità deiprogrammi software necessari per sviluppare i sistemi di intelligenza artificiale. Per esseretecnicamente accurati, però, non è sufficiente avere la disponibilità dei programmi software ma sonoessenziali i dati usati per l'addestramento iniziale, o meglio, i parametri calcolati sulla base diquesti dati. Nel caso dei sistemi attuali, il numero di questi parametri è enor-me, si parla di 175miliardi di parametri (B125) per ChatGPT 3.5, mentre ChatGPT 4 potrebbe averne 100.000 miliardi(T100), come ha affermato Andrew Feldman, CEO di Cerebras, in una conversazione con OpenAI.

Esiste una significativa differenza tra programmi software e parametri. I primi si basano sualgoritmi, spesso pubblicati nella letteratura scientifica specializzata. Negli ultimi anni la scienza

ha fatto enormi passi avanti nella scoperta di algoritmi sempre più sofisticati, basati su complessetecniche matematiche e statistiche. Non serve essere esperti in questi campi per scrivere un programmasoftware a partire da uno di questi algobraio ritmi. Non solo, sul web è possibile trovare tutorial elibrerie che rendono possibile questa attività con un minimo di conoscenze informatiche. Calcolare iparametri è invece un'altra storia. Sono necessari calcolatori molto potenti e un'enorme quantità ditempo di calcolo.

Ad esempio, l'addestramento di ChatGPT 3 è durato 15 giorni, usando in parallelo 10.000 unità dicalcolo, pari a circa quattro milioni di ore di calcolo. Per inciso, una quantità così elevata dicalcolo richiede un consumo energetico che ha effetti ambientali significativi, per non parlare delcosto economico necessario per ottenere tali risorse computazionali.

Un articolo scientifico pubblicato nel feb 2024 da Sayash Kapoor dell'Università di Princeton e daaltri 24 scienziati internazionali ha esaminato l'impatto sulla società dei "modelli fondazionaliaperti". Questi modelli sono impiegati dai sistemi di intelligenza artificiale che usano tecniche diapprendimento automatico basate su grandi quantità di dati. Si tratta di un caso particolare dei"modelli per finalità generali" citati nell'AI Act, in grado di svolgere con competenza un'ampia gammadi compiti distinti, indipendentemente dalle modalità con cui il modello è poi immesso sul mercato.L'articolo mette in discussione l'assunto secondo il quale la pubblica disponibilità dei modellifondazionali favorisce l'avanzamento scientifico del settore e la crescita di un'economia basata sulloro impiego senza significativi effetti collaterali. Al contrario, gli autori evidenziano l'esistenzadi rischi relativi alla sicurezza e alla disinformazione. Rischi dovuti principalmente alladisponibilità di parametri il cui calcolo, come abbiamo già notato, esige ingenti risorsecomputazionali. Si tratta di un patrimonio informativo che consente di ridurre enormemente i costinecessari per la realizzazione di un sistema di intelligenza artificiale, facilitando e anzi favorendochi intende farne un uso malevolo.

Ci troviamo perciò di fronte a un classico dilemma etico: rendere pubblica e facilmente accessibileuna tecnologia che può favorire la crescita della società oppure mantenerla segreta in modo da evitareche sia usata contro la società stessa. Un simile dilemma è sorto per altre tecnologie potenzialmentepericolose, tra cui quella atomica. La tendenza attuale è favorire un'apertura parziale, mantenendosegreti solo alcuni elementi cruciali, i parametri nel caso dei modelli fondazionali. Questa è lastrada seguita da Mistral, anche se il recente ingresso della Microsoft nel capitale sociale dellasocietà francese che ha sviluppato l'omonimo sistema di intelligenza artificiale lascia presagire uncambio di strategia.

Allargando l'orizzonte, è necessario prendere in considerazione il tema della sovranità digitale e, inparticolare, di quella europea. L'entrata in vigore di numerosi regolamenti sul tema dei servizidigitali, dell'accesso ai dati e della protezione della privacy, ha posto le basi per ciò che vieneindicato come il perimetro tecnologico ed economico dell'Unione Europea, una risposta alle politiche avolte aggressive di altri attori internazionali. L'enunciazione e la tutela dei diritti dei cittadinieuropei ha dato finora grande rilievo alla libertà, nelle varie modalità in cui si manifesta: libertàdi movimento, libertà di circolazione dei capitali, e più in generale, libertà delle persone. Lo

sviluppo dei sistemi di intelligenza artificiale solleva numerosi interrogativi nel contestogeopolitico, relativi ai rischi intrinsecamente legati al loro uso. La scienza può, e deve, forniregli strumenti concettuali per consentire alla politica di valutare questi rischi e mettere in attoadeguate strategie di mitigazione.
RIPRODUZIONE RISERVATA Giuseppe Veneziano, “Il pensatore di Novembre”, 2024 / “Art Design”/FabbricaEos.