DB non solo diritto bancario

leggi l'articolo nel sito www.dirittobancario.it

ATTUALITÀ

AI act: bene l'accordo, ma sui foundation model la strada è ancora lunga

8 Gennaio 2024

Massimiliano Masnada, Partner, Hogan Lovells
Valerio Natale, Senior Associate, Hogan Lovells

Massimiliano Masnada, Partner, Hogan Lovells

Valerio Natale, Senior Associate**,** Hogan Lovells

Massimiliano Masnada

Massimiliano Masnada è il socio responsabile del team di Privacy e Cybersecurity di Hogan Lovells in Italia, all'interno del dipartimento SOAR (Strategic Operations, Agreements and Regulation). È uno degli avvocati leader in Italia in materia di privacy e protezione dei dati personali, occupandosi, sin dal 2000, di fornire assistenza sia nella fase stragiudiziale che di contenzioso anche presso il Garante per la protezione dei dati personali.

1. L'accordo raggiunto

Dopo tre giorni di intense discussioni, le istituzioni europee hanno raggiunto un'intesa preliminare per stabilire regole comuni sull'uso dell'IA. Gli elementi principali di questo accordo provvisorio riguardano la definizione di regole per i sistemi di IA ad alto rischio, una revisione del sistema di *governance* con maggiori poteri di enforcement a livello UE, un elenco ampliato di IA vietate o limitate – specialmente con riferimento all'utilizzo delle IA nell'ambito della sicurezza – e, infine, degli obblighi di trasparenza per i cd. *foundation model*, o modelli di fondazione, che dovranno essere soddisfatti prima della loro immissione sul mercato.

2. Cosa è un modello di fondazione

Ma cosa si intende, esattamente, per modello di fondazione, e perché il Parlamento UE ha insistito per prevedere delle regole più stringenti?

Un modello di fondazione è un sistema di IA cui possono ricorrere altri sviluppatori per elaborare nuovi sistemi di IA che utilizzano tale modello di fondazione quale "base" per il proprio sistema. Possiamo immaginare questi modelli di fondazione come grandi libri contenenti conoscenze che gli altri sistemi di IA possono utilizzare per svolgere compiti intelligenti come capire il linguaggio, riconoscere immagini o prendere decisioni, senza dover essere istruiti da zero. La maggior parte dei sistemi di IA, specialmente quelli di IA generativa, infatti, si basano su determinazioni probabilistiche e sono costruiti usando enormi quantità di dati che consentono al modello di imparare a prevedere e costruire elementi ulteriori a partire da un elemento dato. Grazie all'enorme quantità di testo elaborata dal modello di fondazione nella fase di addestramento grazie alle attività di *web scraping*, il modello di fondazione di tipo testuale acquisisce la conoscenza del vocabolario, della sintassi, delle nozioni del mondo e in qualche modo perfino delle capacità di ragionamento. Col tempo, il modello riesce a prevedere come le parole si relazionano tra di esse e quindi a predire, per ogni data parola, la parola successiva più probabile, costruendo così – parola dopo parola – la frase. Ad esempio, a un modello fondato sul testo può essere richiesto di completare la frase "*Invece di guardare un film, Marco ha deciso di leggere un …*". Prima dell'addestramento, il modello non è in grado di comprendere la struttura linguistica e produce quindi parole in maniera casuale, senza coerenza con il contesto della frase. Il modello potrebbe provare a

vedi l'articolo online

completare la frase con la parola "*pennarello*", cioè una parola del tutto decontestualizzata. Tuttavia, durante la fase di apprendimento, grazie all'enorme quantità di testo letto, il modello acquisisce la conoscenza del vocabolario, della sintassi e della relazione esistente tra le parole, e riesce a predire la parola successiva più probabile. Nella frase di esempio, la parola successiva potrebbe essere "*libro*" – "*invece di guardare un film, ha deciso di leggere un libro*". Esistono però anche altre parole adatte a quel contesto, e il risultato potrebbe essere diverso – "*Invece di guardare un film, ha deciso di leggere un giornale*". Dato che tante sono le parole che possono seguire una frase troncata, ogni risposta fornita dal modello preserverà un elemento di casualità, e le risposte alla stessa richiesta potrebbero essere diverse di volta in volta.

È esattamente ciò che ha affascinato il mondo quest'anno, in cui abbiamo sperimentato – diversamente da quanto avviene con un motore di ricerca che si limita a restituire quanto esistente in rete – che formulando la stessa richiesta in un modo o in un altro, o fornendo più o meno contesto, può esserci grande differenza nel risultato, spesso con risultati sorprendentemente creativi.

3. Perché i modelli di fondazione sono così importanti per lo sviluppo di sistemi di IA

I modelli di fondazione sono fondamentali per lo sviluppo dell'IA nei vari campi perché rappresentano una base di conoscenza condivisa che può essere utilizzata da programmatori dotati di meno risorse, che non hanno gli strumenti per sviluppare internamente un modello così approfondito. I modelli di fondazione richiedono infatti investimenti significativi in virtù, da un lato, della necessità di effettuare l'addestramento su enormi quantità di dati (ad oggi, le dimensioni di dati necessarie sono tali da aver spinto gli sviluppatori dei modelli di fondazione a ricorrere all'intero web, che è raccolto tramite attività di *web scraping*) e, dall'altro, della straordinaria potenza di calcolo necessaria ad elaborare un così ampio volume di dati. Successivamente alla prima fase di costruzione principale del modello, questo è poi solitamente affinato mediante delle successive fasi di addestramento basate su *corpus* di dati selezionati dall'uomo e, anche, mediante fasi di addestramento effettuate sulla base di *input* generati da addestratori umani. Dato l'enorme lavoro che esiste dietro l'elaborazione di un modello di fondazione, non è un caso che ad oggi i principali modelli sono quelli sviluppati da grandi realtà, come OpenAI – che ha investito anni in attività di addestramento algoritmico su grande scala prima di divenire celebre – e Google, che da pochi giorni ha annunciato il suo modello di fondazione Gemini.

4. Il rischio più grande nei modelli di fondazione è la presenza di bias nascosti

Tuttavia, se la capacità di questi modelli di fondazione di comprendere e generare dati li rende strumenti utili per lo sviluppo di ulteriori sistemi di IA, che possono essere sviluppati a partire da essi con varie tecniche (come l'ulteriore addestramento su set di dati settoriali e/o proprietari, o l'integrazione con ulteriori tecnologie), allo stesso tempo solleva preoccupazioni riguardo al pericolo che eventuali *bias* di tali modelli di fondazione possano riprodursi a cascata in tutte le applicazioni pratiche di IA.

L'origine di questi *bias* risiede nel fatto che i modelli di fondazione, essendo spesso addestrati su insiemi di dati raccolti con attività di *web scraping*, riproducano tutti quegli elementi di disuguaglianza, pregiudizio o comunque culturali, presenti in tali dati di addestramento. Alcune fonti web possono essere intrinsecamente influenzati da pregiudizi, stereotipi o punti di vista discriminatori, e se tali fonti sono utilizzate alla base dell'addestramento del modello di fondazione, tali pregiudizi non possono che riflettersi nelle risposte del modello, perpetuandone a sua volta gli effetti nelle tecnologie di IA sviluppate a partire da tali modelli.

5. Oltre l'accordo: la necessità di un coordinamento normativo tra AI Act, GDPR e non solo.

È evidente che l'equilibrio tra innovazione tecnologica e protezione dei diritti fondamentali rimane una sfida ancora aperta, non risolvibile con un mero esercizio di trasparenza. Rimangono però ancora da risolvere questioni cruciali legate alla sovrapposizione tra il futuro AI Act e l'attuale GDPR con riferimento ai modelli di fondazione, ad oggi costruiti per lo più sulla base di attività massive di scraping. Rimane ancora aperto, a questo proposito, l'interrogativo circa la corretta base giuridica per la raccolta massiva di dati di tipo particolare sulla rete, stante l'impossibilità di ricorrere al legittimo interesse, se non per ristrettissimi ambiti legati alla ricerca scientifica. Irrisolto è pure il tema delle garanzie poste a tutela dell'esercizio del diritto di opposizione al trattamento, e al momento di esercizio di tale diritto (nella fase di training, nella fase di generazione del modello, o nella fase di output, sottoforma di filtro ai risultati). Ad oggi le autorità di controllo non hanno saputo fornire una risposta coesa alla problematica, confondendo, spesso, tra diritto di cancellazione e diritto di opposizione. Altrettanto problematico sarà il coordinamento tra la disciplina dell'AI Act, come accordata, e la sovrapposizione con gli obblighi di trasparenza già posti dalla normativa consumeristica, su cui vi è peraltro una copiosa giurisprudenza

vedi l'articolo online

sedimentatasi negli anni che dovrò coordinarsi. Infine, rimane tutta da scrivere la disciplina sul regime di responsabilità derivante dall'utilizzo dell'IA, oggetto della proposta di Direttiva UE relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale presentata a fine 2022, su cui il Parlamento UE non si è ancora pronunciato, e che richiederà un coordinamento con l'assetto consolidato della Direttiva E-Commerce e dell'imminente Digital Services Act.

È evidente, dunque, come quest'accordo provvisorio – che peraltro vedrà una sedimentazione normativa soltanto nelle prossime settimane – costituisca l'inizio di una strada ancora molto lunga e ambiziosa. Risolti i problemi legati all'etica e alla trasparenza algoritmica, il punto di arrivo di questa lunga strada da percorrere dipenderà dalla presa di coscienza e dalla capacità del legislatore europeo di coordinare le nuove disposizioni dell'AI Act con le normative già esistenti, su tutte il GDPR, la disciplina a tutela dei consumatori e il regime di responsabilità degli ISP.

vedi l'articolo online

DB non solo diritto bancario

A NEW DIGITAL EXPERIENCE

dirittobancario.it