

Marco Ciurcina

Libertà di ricerca, licenze libere e sistemi d'intelligenza artificiale

(doi: 10.53227/108473)

Rivista di Digital Politics (ISSN 2785-0072)

Fascicolo 2, maggio-agosto 2023

Ente di afferenza:

()

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

Questo articolo è reso disponibile con licenza CC BY NC ND. Per altre informazioni si veda <https://www.rivisteweb.it/>

Marco Ciurcina

Libertà di ricerca, licenze libere e sistemi d'intelligenza artificiale

FREEDOM OF RESEARCH, FREE LICENCES AND ARTIFICIAL INTELLIGENCE SYSTEMS

After years in which many developments in artificial intelligence systems were the outcome of scientific research and were made available under free licences, the recent success of some generative artificial intelligence systems, which were developed using large computing capacities and quantities of data, leads to the question whether conditions are emerging in which a few players will be able to provide artificial intelligence systems, concentrating economic (but not only) power. But the development of artificial intelligence systems continues. Policies are therefore useful that, by exploiting the regulatory framework (favourable to the development of transparent systems and to the scientific research), make data and computing resources available to researchers in order to foster the free availability, study and understanding of artificial intelligence systems. To achieve this, it is certainly useful to make use of the methods already adopted for free software (which, on the one hand, have led to the generation of digital commons and, on the other, are usefully employed by companies). But this is not enough: artificial intelligence systems do not consist of software alone and require computing and data resources. It is supportive of the European regulatory framework which allows it to adopt policies that enhance the scientific research in the development of artificial intelligence systems thus balancing fundamental rights with conflicting objectives: freedom of enterprise, freedom of scientific research and protection of personal data.

KEYWORDS *Artificial Intelligence, Scientific Research, Open Source, Free Software, Law, Policy.*

1. Il problema

Nell'ultimo anno sono stati lanciati diversi servizi di *chatbot* basati su Llm (*Large language model*) trasformativi, un particolare tipo di sistemi d'in-

Mario Ciurcina, StudioLegale.it – Via San Quintino, 32 – 10121 Torino, email: ciurcina@studiolegale.it

telligenza artificiale generativa, tra i quali ChatGpt¹ (basato su Gpt²) di OpenAI e Bard³ (basato su PaLM) (Narang *et al.* 2022) di Google.

Nello stesso periodo sono stati lanciati diversi servizi basati su Llm trasformativi che generano immagini a partire da descrizioni testuali (tra i quali Stable Diffusion⁴ di Stability AI, Dall-e⁵ di OpenAI e Midjourney⁶ di Midjourney Inc.).

Si è anche prodotto un forte interesse mediatico ai sistemi d'intelligenza artificiale, in particolare a quelli generativi come gli Llm trasformativi, producendo nell'opinione pubblica la percezione che tali sistemi siano destinati a trovare spazio nella vita di tutti in tempi molto rapidi.

Lo sviluppo di Gpt e Palm ha richiesto l'uso di grandi capacità di calcolo e quantità di dati e quindi s'ipotizza che si possano generare condizioni di oligopolio nella fornitura dei servizi basati su Llm trasformativi.

I (pochi) fornitori dei servizi abilitati dai sistemi d'intelligenza artificiale generativa hanno interesse a proteggere la loro posizione di vantaggio nel mercato. Per massimizzare il profitto che se ne ricava, le imprese possono trovare utile non divulgare le informazioni che consentono di studiare, comprendere, riprodurre e verificare il funzionamento di un sistema d'intelligenza artificiale. Il segreto industriale, gli altri diritti d'esclusiva e le misure tecniche adottate con lo scopo di ostacolare la possibilità di studiare, comprendere, riprodurre e verificare il funzionamento di un sistema d'intelligenza artificiale portano l'attività di sviluppo di tale sistema lontano dalla scienza ma attuano il diritto alla libertà d'impresa, riconosciuto dall'art. 16 della Carta dei diritti fondamentali dell'Unione europea⁷.

Condizioni di mercato strutturalmente concentrato, come ricorda il presidente Mattarella nel discorso tenuto il 15 settembre 2023 all'Assemblea generale di Confindustria⁸, non sono compatibili con la democrazia: «Qual è un principio fondamentale della democrazia? Evitare la concentrazione del potere, a garanzia della libertà di tutti. Vale per le istituzioni e per le imprese. E la lotta ai monopoli ne rappresenta un capitolo importante. Vanno rifiutate spinte di ingiustificate egemonie delle istituzioni nella gestione delle regole o, all'oppo-

¹ Vedi, <https://chat.openai.com/chat>.

² Vedi, <https://openai.com/gpt-4>.

³ Vedi, <https://bard.google.com>.

⁴ Vedi, <https://github.com/Stability-AI/stablediffusion>.

⁵ Vedi, <https://labs.openai.com/>.

⁶ Vedi, <https://www.midjourney.com/>.

⁷ Che dispone: «È riconosciuta la libertà d'impresa, conformemente al diritto comunitario e alle legislazioni e prassi nazionali». Vedi, [https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000X1218\(01\):IT:HTML](https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000X1218(01):IT:HTML).

⁸ Vedi, <https://www.quirinale.it/elementi/98056>.

sto, di pseudo-assolutismo imprenditoriale, magari veicolato dai nuovi giganti degli «over the top» che si pretendono, spesso, *legibus soluti* (Carugati 2023).

Se pochi operatori potranno fornire sistemi d'intelligenza artificiale generativa utilizzabili come servizio, concentreranno un rilevante potere economico (ma non solo).

È quindi utile porsi una domanda: è ragionevole ipotizzare che l'esito della concentrazione di mercato nella fornitura dei sistemi d'intelligenza artificiale generativa (o, in senso più ampio, dei sistemi d'intelligenza artificiale) sia consolidato ed inevitabile? Che politiche si possono adottare, nel contesto attuale, per favorire lo sviluppo di sistemi d'intelligenza artificiale (efficienti ed efficaci ma anche) sostenibili, rispettosi dei diritti fondamentali e che favoriscono la distribuzione delle risorse, la riduzione delle diseguaglianze e il bene comune?

Per rispondere alle domande è utile analizzare il contesto normativo, in particolare le norme che incentivano lo sviluppo di sistemi trasparenti e realizzati nel quadro di attività di ricerca scientifica, e quindi valutare in che modo tali incentivi impattano sullo sviluppo dei sistemi d'intelligenza artificiale.

È quindi utile considerare l'uso delle licenze libere nella pratica delle imprese e dei centri di ricerca sull'intelligenza artificiale (che continua anche negli ultimi anni) e il rapporto tra licenze libere e ricerca scientifica.

Si osserva che l'uso delle licenze libere è utile per incentivare la ricerca scientifica, ma non è sufficiente.

Si indicano infine gli spazi di opportunità offerti dal quadro normativo vigente, che permettono di adottare politiche di favore per la ricerca scientifica sui sistemi d'intelligenza artificiale bilanciando i diritti fondamentali che, negli ultimi anni, hanno assunto un ruolo centrale nell'attività di normazione dell'Unione Europea, in particolare nel digitale (De Gregorio 2021).

2. Il contesto normativo

Incentivi alla trasparenza dei sistemi d'intelligenza artificiale

Diverse norme (in vigore o in corso di approvazione) richiedono che i sistemi d'intelligenza artificiale siano trasparenti.

La proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate in materia di intelligenza artificiale (legge sull'intelligenza artificiale) e che modifica taluni atti legislativi dell'Unione del 21 aprile 2021⁹ nella versione con gli emendamenti approvati dal Parlamento

⁹ Vedi, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex%3A52021PC0206>.

europeo il 14 giugno 2023¹⁰, tratta la trasparenza dei sistemi in diverse disposizioni. La trasparenza sul funzionamento dei sistemi per gli utenti è disciplinata all'art. 13 (Trasparenza e fornitura di informazioni) il quale richiede che i sistemi d'intelligenza artificiale ad alto rischio siano progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire ai fornitori e agli utenti di comprendere ragionevolmente il funzionamento del sistema.

Inoltre, l'art. 11 prevede che il fornitore di un sistema d'intelligenza artificiale ad alto rischio rediga idonea documentazione tecnica del sistema.

Ai sensi dell'art. 50 il fornitore tiene la documentazione tecnica a disposizione delle autorità nazionali competenti.

Ma ai sensi dell'art. 70 sia le autorità nazionali competenti che gli organismi notificati sono tenuti a rispettare la riservatezza delle informazioni e dei dati ottenuti nello svolgimento dei loro compiti e delle loro attività in modo da tutelare i diritti di proprietà intellettuale e le informazioni commerciali riservate o i segreti commerciali, compreso il codice sorgente.

Gli obblighi previsti per i cd. «foundation models» (cioè i modelli d'intelligenza artificiale che all'art. 3, lett. c) sono definiti come quelli che sono addestrati su dati d'ampia scala, sono progettati per la generalità dei risultati e possono essere adattati a un'ampia gamma di distinti compiti specifici) sono previsti agli artt. 28 e 28 b).

In sintesi, la proposta di legge sull'intelligenza artificiale prevede obblighi di ampia trasparenza solo in ambiti circoscritti (per alcuni dei sistemi ad alto rischio) e, comunque, tenendo ferma la tutela dei diritti di riservatezza delle imprese in modi non necessariamente compatibili con la ricerca scientifica.

Il problema della trasparenza dei sistemi d'intelligenza artificiale era già emerso prima della pubblicazione della proposta di legge sull'intelligenza artificiale.

All'art. 13.2, lett. f) e all'art. 14.2, lett. g) del Regolamento Ue 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati)¹¹ si prevede che il titolare del trattamento deve fornire all'interessato le seguenti informazioni: «l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale

¹⁰ Vedi, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_IT.html.

¹¹ Vedi, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32016R0679>.

trattamento per l'interessato.». La formulazione di queste norme ha aperto un dibattito (che continua) su come debba essere soddisfatto l'obbligo di fornire «informazioni significative sulla logica utilizzata» quando i processi decisionali automatizzati siano realizzati facendo uso di sistemi d'intelligenza artificiale.

Un altro limite all'adozione di sistemi d'intelligenza artificiale «oscuri» potrebbe darsi quando si utilizzino tali sistemi nel quadro di procedimenti della pubblica amministrazione. Anche se sul tema il dibattito resta aperto e ci sono pronunce divergenti, in Italia il Tar del Lazio (sentenze 9224/2018, 9230/2018, 2936/2019, 8474/2019, 7370/2020, 7526/2020 e 13692/2020) e il Consiglio di Stato (sentenze 2270/2019, 2936/2019, 8472/2019 e 8474/2019, 881/2020) hanno avuto modo di statuire che le norme sul diritto d'accesso e sulla trasparenza della pubblica amministrazione implicano il diritto di accedere alla regola informatica, ossia l'algoritmo, attraverso cui si giunge alla decisione amministrativa. Sulla base di questo orientamento, quando la decisione amministrativa sia anche solo parzialmente basata su un sistema d'intelligenza artificiale, ci si può domandare come vada soddisfatto quest'obbligo.

Incentivi alla ricerca scientifica sui sistemi d'intelligenza artificiale

La libertà di ricerca scientifica sui sistemi d'intelligenza artificiale (come in qualsiasi altro ambito) è riconosciuta dall'art. 13 della Carta dei diritti fondamentali dell'Unione europea¹².

L'Unione europea adotta una politica di favore per la scienza aperta (Commissione europea 2019)¹³ che si concreta anche in specifici interventi normativi.

Per esempio, la direttiva 2019/1024/Ue del Parlamento europeo e del Consiglio del 20 giugno 2019 relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico («Open Data directive» o «Direttiva dati aperti»)¹⁴, all'art. 10, prevede che gli Stati membri promuovono la disponibilità dei dati della ricerca adottando politiche nazionali e azioni pertinenti per rendere i dati della ricerca finanziata con fondi pubblici apertamente disponibili. La norma è stata recepita in Italia all'art. 9-bis («Dati della ricerca») del Dlgs n. 36/2006 del 24 gennaio, attuazione della direttiva Ue 2019/1024

¹² Che dispone: «le arti e la ricerca scientifica sono libere. La libertà accademica è rispettata». Vedi, [https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000X1218\(01\):IT:HTML](https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000X1218(01):IT:HTML).

¹³ Vedi, https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en.

¹⁴ Vedi, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32019L1024>.

relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico che ha abrogato la direttiva 2003/98/Ce.

Di particolare interesse nel settore dell'intelligenza artificiale è l'art. 3 («Estrazione di testo e di dati per scopi di ricerca scientifica») della direttiva 790/2019/UE del Parlamento europeo e del Consiglio del 17 aprile 2019 sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/Ce e 2001/29/Ce¹⁵ il quale prevede un'eccezione espressa al diritto d'autore e al diritto *sui generis* del costituente di banche di dati per le attività di estrazione di testo e di dati da opere o altri materiali utilizzati per scopi di ricerca scientifica. In Italia l'eccezione è stata recepita con l'adozione del nuovo art. 70-ter della Legge n. 633/1941 del 22 aprile («Protezione del diritto d'autore e di altri diritti connessi al suo esercizio»)¹⁶. Tale previsione riduce il rischio di vertenze giudiziali per chi sviluppi e distribuisca sistemi d'intelligenza artificiale per finalità di ricerca scientifica, rischio che, invece, è concreto per chi li sviluppi e distribuisca per finalità diverse¹⁷.

Anche la proposta di legge sull'intelligenza artificiale¹⁸, nella versione con gli emendamenti approvati dal Parlamento europeo il 14 giugno 2023¹⁹ favorisce la ricerca scientifica: da una parte, prevede obblighi stringenti per chi procede ad immettere sul mercato o mettere in servizio sistemi d'intelligenza artificiale, dall'altra, all'art. 2, paragrafo 5-quinquies, prevede espressamente che «il presente regolamento non si applica alle attività di ricerca, prova e sviluppo relative ai sistemi di Ia prima che tali sistemi siano immessi sul mercato o messi in servizio, a condizione che tali attività siano condotte nel rispetto dei diritti fondamentali e della normativa dell'Unione applicabile».

Anche il Regolamento generale sulla protezione dei dati, che attua il diritto alla protezione dei dati personali previsto dall'art. 8 della Carta dei diritti fondamentali dell'Unione europea²⁰, prevede alcune semplificazioni per l'at-

¹⁵ Vedi, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:32019L0790>.

¹⁶ Vedi, <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1941-04-22;633!vig=2023-08-06>.

¹⁷ Vedi, <https://www.thefashionlaw.com/from-chatgpt-to-deepfake-creating-apps-a-running-list-of-key-ai-lawsuits/>, <https://www.ft.com/content/704d0bba-2653-4a27-bee1-ec45c6ed1080>.

¹⁸ Vedi nota 11.

¹⁹ Vedi nota 12.

²⁰ Che dispone: «1. Ogni individuo ha diritto alla protezione dei dati di carattere personale che lo riguardano. 2. Tali dati devono essere trattati secondo il principio di lealtà, per finalità determinate e in base al consenso della persona interessata o a un altro fondamento legittimo previsto dalla legge. Ogni individuo ha il diritto di accedere ai dati raccolti che lo riguardano e di ottenerne la rettifica. 3. Il rispetto di tali regole è soggetto al controllo di un'autorità indipendente». Vedi, [https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000X1218\(01\):IT:HTML](https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000X1218(01):IT:HTML).

tività di ricerca scientifica (per esempio, eccezioni ai principi di limitazione della finalità, di limitazione della conservazione e di trasparenza, le condizioni da soddisfare per trattare lecitamente i dati personali in mancanza di consenso, eccezione al diritto di cancellazione²¹) bilanciando il diritto alla tutela dei dati personali con la libertà di ricerca scientifica.

3. Sistemi d'intelligenza artificiale tra trasparenza e ricerca scientifica

L'analisi del contesto normativo porta a concludere che ci sono norme che favoriscono lo sviluppo di sistemi d'intelligenza trasparente e/o realizzati nello svolgimento di attività di ricerca scientifica.

È utile qualche considerazione sul rapporto tra i due aspetti.

La più ampia trasparenza di un sistema d'intelligenza artificiale implica la possibilità di usarlo, riprodurlo e verificarne il funzionamento.

Adottando un approccio ripetibile e documentandolo in maniera adeguata è possibile realizzare sistemi d'intelligenza artificiale riproducibili e verificabili (Gundersen *et al.* 2018a), ma gli studi di settore soddisfano solo in piccola parte i requisiti di riproducibilità, anche se si osserva un trend di crescita nel tempo (Gundersen *et al.* 2018b; Heaven 2020).

Alcuni studiosi evidenziano che per riprodurre un sistema d'intelligenza artificiale ad apprendimento automatico è necessario adottare un approccio ripetibile al lavoro di creazione del sistema e documentare precisamente ciò che viene fatto per realizzarlo (Gundersen *et al.* 2018a; Carter *et al.* 2019). Altri studiosi domandano che i criteri di riproducibilità e verificabilità siano tenuti in considerazione (Haibe-Kains *et al.* 2020).

Ma la riproducibilità dei sistemi d'intelligenza artificiale può essere intesa in modi diversi; per esempio, si evidenzia l'utilità di distinguere tra riproducibilità dei metodi, riproducibilità dei risultati e riproducibilità inferenziale (Gundersen *et al.* 2018b). Inoltre, l'esatta riproduzione del funzionamento dei sistemi d'intelligenza artificiale non è conseguibile con certe configurazioni *hardware* (Carter 2019).

²¹ Vedi art. 5.1, lett. b), art. 5.1, lett. e), art. 14.5, lett. b), art. 9.2, lett. j) e art. 17.3, lett. d). È però anche vero che il Regolamento generale sulla protezione dei dati all'art. 9.4 prevede che «gli Stati membri possono mantenere o introdurre ulteriori condizioni, comprese limitazioni, con riguardo al trattamento di dati genetici, dati biometrici o dati relativi alla salute». L'Italia si è avvalsa di questa facoltà adottando gli artt. 2-*quater* («Regole deontologiche») e 2-*septies* («Misure di garanzia per il trattamento dei dati genetici, biometrici e relativi alla salute») del Dlgs n. 196/2003, ma, ad oggi, il quadro regolatorio non è ancora definitivo.

Quali criteri soddisfino il metodo scientifico è quindi oggetto di discussione nella comunità scientifica.

A fronte di ciò, alcuni autori, pur prendendo atto del problema, raccomandano che i sistemi d'intelligenza artificiale siano almeno «spiegabili» (Floridi *et al.* 2018).

D'altronde, autorevoli studiosi illustrano percorsi di sviluppo della ricerca scientifica che permetteranno di acquisire nuove capacità di controllo e comprensione e quindi permetteranno di sviluppare sistemi d'intelligenza artificiale sicuri, trasparenti e verificabili (Tegmark *et al.* 2023).

Il rapporto tra trasparenza e ricerca scientifica è certamente un rapporto complesso che è destinato ad evolvere nel tempo.

Ma al fine di individuare quali politiche adottare per incentivare la ricerca scientifica, non è necessario approfondire se, come e quali aspetti dei sistemi d'intelligenza artificiale debbano essere riproducibili per soddisfare il metodo scientifico (la soluzione dipende anche da cosa s'intende per metodo scientifico), né se la riproducibilità dei sistemi d'intelligenza artificiale che oggi non sono riproducibili sarà in futuro superabile.

È invece utile individuare percorsi che favoriscano la ricerca scientifica su sistemi d'intelligenza artificiale liberamente disponibili che si possano studiare, comprendere, riprodurre e verificare e che quindi favoriscano la generazione di beni comuni digitali, come mettono in luce le considerazioni che seguono.

4. Licenze libere e beni comuni digitali

Per il software, la riproducibilità dipende dalla possibilità di accedere al codice sorgente e alle istruzioni e agli strumenti necessari per riprodurre la forma utile al funzionamento dello stesso software.

L'accesso a tali elementi (il codice sorgente completo del software) è un requisito della definizione di software libero: il software che è disponibile secondo i termini di licenze che permettono che l'utente abbia la libertà di usare, modificare e distribuire, anche in versione modificata, lo stesso software (Ciurcina 2017).

Il software libero è quindi riproducibile e verificabile; anche per questo è utile a favorire comportamenti responsabili nei soggetti che lo sviluppano, e quindi a generare fiducia negli utenti.

Il software libero nasce negli anni Ottanta e si diffonde in comunità di persone che condividono i valori della libertà d'informazione e di conoscenza, figli della cultura accademica e scientifica (Ciurcina 2020).

Negli ultimi decenni le licenze di software libero si sono dimostrate utili a catalizzare la generazione di beni comuni digitali (come per esempio il «Ker-

nel linux»): risorse a disposizione di tutti alle quali contribuiscono molti soggetti. Nello stesso periodo il software libero è stato adottato in misura crescente dalle imprese, che hanno saputo raffinare modelli di attività economica idonei a sfruttare il software libero per fornire prodotti e servizi (Berlinguer 2020; Ciurcina 2020).

Il modello delle licenze libere si è dimostrato utile anche per generare beni comuni digitali che non sono costituiti da software, come per esempio Wikipedia, l'enciclopedia online collaborativa, e Open street map, la mappa online collaborativa (Bassi *et al.* 2018).

Ma, ci si domanda, è sufficiente distribuire i sistemi d'intelligenza artificiale con delle licenze libere per incentivare la generazione di beni comuni digitali?

5. Sistemi d'intelligenza artificiale e licenze libere

Al fine d'individuare politiche utili ad incentivare la generazione di beni comuni digitali, sono utili alcune considerazioni sull'attuale sviluppo dei sistemi d'intelligenza artificiale.

Prima che i sistemi d'intelligenza artificiale generativi s'imponessero all'attenzione del grande pubblico, l'inclinazione alla produzione di condizioni oligopolistiche era tutt'altro che scontata; è quindi utile dar conto dei trend che, ancora oggi, muovono in direzioni divergenti.

I più utilizzati framework di sviluppo di sistemi d'intelligenza artificiale sono oggi disponibili con licenza libera: per esempio, «TensorFlow» è distribuito con licenza Apache v. 2.0 e «Pytorch» con licenza Bsd²². Anche altri strumenti di sviluppo di sistemi d'intelligenza artificiale sono disponibili con licenza libera; alcuni di questi sono elencati tra i progetti della Lf Ai & Data foundation, un'iniziativa della Linux foundation alla quale partecipano numerose imprese attive del settore che ha come obiettivo quello di raccogliere e gestire fondi a sostegno di «progetti *open source* legati all'intelligenza artificiale, all'apprendimento automatico e ai dati»²³.

Anche dei sistemi d'intelligenza artificiale calibrati/addestrati sono disponibili con licenza libera. Per esempio, il sistema «Gpt-2»²⁴ è distribuito con licenza Mit da OpenAI. È interessante osservare che la stessa OpenAI ha successivamente realizzato nuove versioni del sistema Gpt, la più recente delle quali è la versione «Gpt-4», che non sono state distribuite con licenza libera.

²² Vedi, https://en.wikipedia.org/wiki/Comparison_of_deep-learning_software.

²³ Vedi, <https://lfaidata.foundation/>.

²⁴ Vedi, <https://github.com/openai/gpt-2>.

Recentemente, Meta ha distribuito il Llm «Llama2»²⁵ secondo i termini della Llama 2 community license agreement²⁶, che consente di «usare, riprodurre, distribuire, copiare, creare opere derivate e apportare modifiche», ma non è stata considerata compatibile con la definizione di «*open source*»²⁷ perchè prevede alcune limitazioni d'uso.

Inoltre, il Technology innovation institute²⁸ ha distribuito il «Llm Falcon 40B»²⁹ secondo i termini della licenza Apache v. 2.0 e il «Llm Falcon 180B»³⁰ secondo i termini della «Falcon-180B TII license»³¹ (che è basata sulla licenza Apache v. 2.0, ma prevede alcune limitazioni d'uso).

Nel progetto di sviluppo del «Llm Bloom»³² (frutto del lavoro collaborativo di più di 1.000 ricercatori) si adotta la nuova licenza Responsible ai license (Rail)³³, che però non soddisfa le definizioni di software libero e di *open source*.

In sintesi, le licenze libere (ma anche licenze simili a quelle, che introducono alcune limitazioni d'uso) sono state utilizzate, e continuano ad esserlo, da ricercatori e imprese che sviluppano sistemi d'intelligenza artificiale: le licenze libere soddisfano l'interesse di entrambe le comunità (Engler 2021) e, secondo alcuni, la disponibilità con licenza libera dei sistemi d'intelligenza artificiale fa progredire i progetti più rapidamente di quelli non disponibili liberamente (Patel e Ahmad 2023).

È interessante osservare che la proposta di legge sull'intelligenza artificiale³⁴ nella versione con gli emendamenti approvati dal Parlamento europeo il 14 giugno 2023³⁵, coglie l'importanza di valorizzare il modello di licenza del software libero e *open source*. Mentre nella proposta originaria della Commissione l'*open source* non era considerato, l'attuale formulazione tratta del tema ai considerando 12 lett. a), 12 lett. b), 12 lett. c) e 60, all'art. 2 paragrafo 5 lett. e), e all'art. 28 lett. b). L'art. 5 lett. e) prevede espressamente un'eccezione per i componenti di intelligenza artificiale disponibili secondo i termini di licenze *open source*, salvo che si tratti di *foundation models*.

²⁵ Vedi, <https://ai.meta.com/llama/>, <https://github.com/facebookresearch/llama>.

²⁶ Vedi, <https://github.com/facebookresearch/llama/blob/main/LICENSE>.

²⁷ Vedi, <https://blog.opensource.org/metals-llama-2-license-is-not-open-source/>.

²⁸ Vedi, <https://www.tii.ae/>.

²⁹ Vedi, <https://huggingface.co/tiiuae/falcon-40b>.

³⁰ Vedi, <https://huggingface.co/tiiuae/falcon-180b>.

³¹ Vedi, <https://huggingface.co/spaces/tiiuae/falcon-180b-license/blob/main/LICENSE.txt>.

³² Vedi, <https://huggingface.co/bigscience/bloom>.

³³ Vedi, <https://huggingface.co/spaces/bigscience/license>.

³⁴ Vedi nota 11.

³⁵ Vedi nota 12.

6. Sistemi d'intelligenza artificiale tra licenze libere e ricerca scientifica

La disponibilità con licenza libera del modello calibrato/addestrato e degli strumenti di sviluppo utilizzati per la calibrazione/addestramento non massimizza la possibilità di studiare, comprendere, riprodurre e verificare i sistemi d'intelligenza artificiale: è necessario che siano liberamente disponibili anche tutte le informazioni utili per realizzarli (oltre ad idonee capacità di calcolo). Per esempio, è necessario conoscere l'architettura della rete calibrata/addestrata e la versione degli strumenti software, il protocollo e la base di dati utilizzati per la calibrazione/addestramento.

Peraltro, è controverso se sia sempre corretto utilizzare una licenza libera per un modello calibrato/addestrato: è diffusa l'opinione che, per lo meno negli Usa, il diritto d'autore non tuteli i modelli in senso stretto (Lindberg 2023).

In Europa ci si può domandare se, a seconda dei casi, ai modelli calibrati/addestrati si applichi (oppure no) il diritto d'autore o il diritto *sui generis* del costituente di banca di dati.

È utile ricordare che le licenze libere disponibili generalmente gestiscono il diritto d'autore e il diritto sui brevetti d'invenzione e che solo alcune licenze gestiscono altri diritti simili come i diritti connessi al diritto d'autore e il diritto *sui generis* del costituente di banca di dati.

Quali siano le licenze più idonee a massimizzare la possibilità che dai sistemi d'intelligenza artificiale si generino dei beni comuni digitali (risorse a disposizione di tutti alle quali contribuiscono molti soggetti) resta un tema aperto.

In ogni caso, al fine di favorire la ricerca scientifica e massimizzare la possibilità di studiare, comprendere, riprodurre e verificare i sistemi di intelligenza artificiale, dovrebbero essere liberamente disponibili tutte le informazioni utili, compresi i dati utilizzati per la calibrazione/addestramento.

Per esempio, il progetto «Llm Bloom»³⁶, che adotta un modello aperto e collaborativo di sviluppo, rende anche disponibili informazioni dettagliate sui dati utilizzati per la calibrazione/addestramento, rendendo più facile studiarlo e comprenderne il funzionamento e quindi eliminare *bias* ed altri difetti.

Ma se anche i dati utilizzati per la calibrazione/addestramento sono disponibili, non vengono meno gli obblighi imposti a protezione dei dati personali delle persone interessate; tra questi, l'obbligo di minimizzare il trattamento dei dati personali e quindi la loro diffusione: se i dati utilizzati per la calibrazione/addestramento dei sistemi d'intelligenza artificiale includono dati personali, la scelta del modo in cui rendere disponibili tali dati deve tenerne

³⁶ Vedi nota 36.

conto. Peraltro, le licenze libere non gestiscono gli obblighi e i diritti previsti dalle norme a protezione dei dati personali.

7. Sistemi di intelligenza artificiale e beni comuni digitali

Ci sono buone ragioni per pensare che lo sviluppo dell'intelligenza artificiale non produrrà inevitabilmente condizioni di oligopolio.

Sono in corso percorsi di ricerca scientifica sui sistemi d'intelligenza artificiale che fanno ipotizzare importanti cambiamenti nel futuro.

Gli incentivi normativi analizzati creano condizioni favorevoli alla ricerca scientifica e allo sviluppo di sistemi d'intelligenza artificiale trasparenti.

È quindi utile adottare politiche che incentivino lo sviluppo di sistemi d'intelligenza artificiale investendo per permettere ai ricercatori che rendono disponibili i risultati del loro lavoro con licenza libera di disporre di ciò che gli occorre: capacità di calcolo e basi di dati.

Un'opportunità è offerta dalle norme adottate dall'Unione europea in attuazione della strategia europea in materia di dati³⁷ In particolare, è interessante l'istituzione delle organizzazioni per l'altruismo dei dati, disciplinate dagli artt. 16 e seguenti del Regolamento Ue n. 868/2022 del Parlamento europeo e del Consiglio del 30 maggio 2022 relativo alla governance europea dei dati e che modifica il Regolamento Ue n. 1724/2018 («Regolamento sulla governance dei dati»)³⁸. Per altruismo dei dati si intende «la condivisione volontaria di dati sulla base del consenso accordato dagli interessati al trattamento dei dati personali che li riguardano, o sulle autorizzazioni di altri titolari dei dati volte a consentire l'uso dei loro dati non personali, senza la richiesta o la ricezione di un compenso che vada oltre la compensazione dei costi sostenuti per mettere a disposizione i propri dati, per obiettivi di interesse generale, stabiliti nel diritto nazionale, ove applicabile, quali... ..la ricerca scientifica nell'interesse generale» (art. 3, n. 10 del Regolamento sulla governance dei dati). L'art. 16 del Regolamento sulla governance dei dati prevede che gli Stati membri possono adottare disposizioni e stabilire politiche nazionali per favorire l'altruismo dei dati. Per esempio, con norme nazionali si potrebbero prevedere:

- investimenti per dotare le organizzazioni per l'altruismo dei dati delle capacità di calcolo e dei dati necessari per consentire ai ricer-

³⁷ Vedi, <https://digital-strategy.ec.europa.eu/it/policies/strategy-data>, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_it.

³⁸ Vedi, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=celex%3A32022R0868>.

catori che rendono disponibili i risultati del loro lavoro con licenza libera di sviluppare dei sistemi d'intelligenza artificiale;

- condizioni di favore (incluse le tariffe, ai sensi dell'art. 6 del Regolamento sulla governance dei dati) per consentire alle organizzazioni per l'altruismo dei dati di accedere ai dati dalla pubblica amministrazione per metterli a disposizione dei ricercatori.

L'art. 17 prevede l'istituzione del registro delle organizzazioni per l'altruismo dei dati riconosciute; le norme successive individuano le caratteristiche di tali organizzazioni e gli obblighi che devono soddisfare per ottenere la registrazione.

L'art. 38, paragrafo 2, del Regolamento sulla governance dei dati prevede che lo stesso si applica dal 24 settembre 2023.

In Italia, l'attuazione delle norme del Regolamento sulla governance dei dati è ancora in corso: l'art. 13 (Delega al governo per l'adeguamento della normativa nazionale alle disposizioni del Regolamento Ue n. 868/2022, relativo alla governance europea dei dati e che modifica il Regolamento Ue n. 1724/2018) della proposta di legge n. 1342 «Delega al governo per il recepimento delle direttive europee e l'attuazione di altri atti dell'Unione europea – Legge di delegazione europea 2022-2023»³⁹ delega il governo ad adottare (entro quattro mesi) dei decreti legislativi per attuare le disposizioni del Regolamento sulla governance dei dati.

Le organizzazioni per l'altruismo dei dati potrebbero quindi favorire lo sviluppo da parte della comunità scientifica di sistemi d'intelligenza artificiale che si possono studiare, comprendere, riprodurre e verificare mettendo a disposizione capacità di calcolo e dati.

Quando la calibrazione/addestramento di sistemi d'intelligenza artificiale richiede l'uso di dati personali, le organizzazioni per l'altruismo dei dati potrebbero garantire che tali dati siano utilizzati nel rispetto dei diritti delle persone interessate. Per esempio, potrebbero far eseguire la calibrazione/addestramento dei sistemi di intelligenza artificiale e le attività di riproduzione e verifica degli stessi nei propri centri di calcolo (e quindi non trasferire i dati personali all'esterno). In questo modo, da una parte si centralizzerebbe l'investimento necessario per realizzare un sistema idoneo a calibrare/addestrare sistemi d'intelligenza artificiale, dall'altra si potrebbe fornire un servizio che protegge i dati personali dei cittadini (rendendo più facile acquisirne la fiducia) e sviluppa sistemi d'intelligenza artificiale disponibili liberamente, con tutte le informazioni utili per studiarli, comprenderli, riprodurli e verificarli.

³⁹ Vedi, <https://www.camera.it/leg19/126?leg=19&idDocumento=1342>.

Consentendo alla comunità scientifica di accedere a tali informazioni si creano le condizioni perché si diffonda fiducia nei sistemi d'intelligenza artificiale.

Assumendo impegno e garantendo che i dati personali siano usati solo per sviluppare (e studiare, comprendere, riprodurre e verificare) sistemi d'intelligenza artificiale disponibili liberamente (Bassi *et al.* 2019) e che siano sostenibili, rispettosi dei diritti fondamentali ed etici (e richiedendo di assumere lo stesso impegno ai ricercatori), le organizzazioni per l'altruismo dei dati potrebbero ottenere più facilmente la fiducia delle persone interessate (che consentono all'uso dei loro dati personali).

Altro quadro normativo interessante è la proposta di Regolamento del Parlamento europeo e del Consiglio riguardante norme armonizzate sull'accesso equo ai dati e sul loro utilizzo (Com/2022/68 final⁴⁰, di cui è stata recentemente annunciata la versione finale concordata da Consiglio e Parlamento europeo⁴¹) che include norme che permettono agli utenti di dispositivi connessi di accedere ai dati generati da tali dispositivi e dai servizi a essi collegati e di condividerli con terzi (incluse le organizzazioni per l'altruismo dei dati). Ciò può contribuire a creare le condizioni perché, disegnando tecnologie e politiche idonee, molte persone siano incentivate ad acquisire i loro dati e metterli a disposizione di un'organizzazione per l'altruismo dei dati.

Favorendo lo sviluppo di sistemi d'intelligenza artificiale che è possibile studiare, comprendere, riprodurre e verificare si incentivano le imprese ad adottare nuovi modelli di attività imprenditoriale che possono contribuire allo sviluppo di sistemi d'intelligenza artificiale che generano beni comuni digitali, bilanciando libertà d'impresa, libertà di ricerca scientifica e protezione dei dati personali.

8. Conclusione

Le imprese hanno saputo riconfigurare i propri modelli di attività economica adattandosi al modello di licenza del software libero e, di conseguenza, hanno contribuito allo sviluppo di beni comuni digitali.

Nello sviluppo dei sistemi d'intelligenza artificiale l'uso delle licenze libere è utile per favorire la ricerca scientifica e la generazione di beni comuni digitali, ma non è sufficiente: servono anche grandi capacità di calcolo e quantità di dati.

Considerando il contesto normativo, che favorisce sistemi trasparenti e la ricerca scientifica, si sono individuate le scelte politiche che permettono di

⁴⁰ Vedi, <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=COM%3A2022%3A68%3AFIN>.

⁴¹ Vedi, <https://www.consilium.europa.eu/it/press/press-releases/2023/06/27/data-act-council-and-parliament-strike-a-deal-on-fair-access-to-and-use-of-data/>.

favorire lo sviluppo di sistemi d'intelligenza artificiale che è possibile studiare, comprendere, riprodurre e verificare; in particolare, si può valorizzare il ruolo delle organizzazioni per l'altruismo dei dati e prevedere: investimenti per dotarle delle capacità di calcolo e dei dati necessari perché i ricercatori che rendono disponibili i risultati del loro lavoro con licenza libera possano sviluppare dei sistemi d'intelligenza artificiale; condizioni di favore (incluse le tariffe) per consentire loro di accedere ai dati dalla pubblica amministrazione per metterli a disposizione dei ricercatori.

Riferimenti bibliografici

- BASSI, E., CIURCINA, M., DE MARTIN, J.C., FENOGLIETTO, S., ROCCHI, G., SAGARRA PASCUA, O. e BRIA F. (2017), *Legal Frameworks for Digital Commons DECODE OS and Legal Guidelines*, Progetto Decode, <https://decodeproject.eu/file/189/download>.
- BASSI, E., CIURCINA, M., DE MARTIN, J.C., e FENOGLIETTO S. (2019), *Licensing of Digital Commons Including Personal Data – Update*, Progetto Decode, <https://decodeproject.eu/file/507/download>.
- BERLINGUER, M. (2020), *Commons, Markets and Public Policy*, in «Transform! Epaper», https://www.transform-network.net/fileadmin/user_upload/2020-01-commons_3.pdf. Consultato il 15 giugno 2023.
- CARTER, R. E., ATTIA, Z. I., LOPEZ-JIMENEZ, F. e FRIEDMAN, P. A. (2019), *Pragmatic Considerations for Fostering Reproducible Research in Artificial Intelligence*, in «Npj Digital Medicine», 2(1), pp. 40–42.
- CARUGATI, A. *Mattarella agli industriali: «No al capitalismo di rapina»*, in «Il Manifesto», 22 settembre, <https://ilmanifesto.it/mattarella-agli-industriali-no-al-capitalismo-di-rapina>.
- CASO, R. (2022), *Data, ricerca scientifica e privatizzazione della conoscenza*, in «Trento Law and Technology», Research Group Research Paper n. 48, <https://zenodo.org/record/5902766>, do, <https://doi.org/10.5281/zenodo.5894533>. Consultato il 15 giugno 2023.
- CIURCINA, M. (2020), *Etica Hacker?*, in «Digitcult. Scientific Journal on Digital Cultures», 5(1), pp. 67-76.
- CIURCINA, M. (2017), *Licenze di software libero ed altre licenze libere: codice genetico di beni comuni digitali*, in «Digitcult. Scientific Journal on Digital Cultures», 2(1), pp. 15-24.
- COMMISSIONE EUROPEA (2019), *Open Science*, Commissione Europea, https://research-and-innovation.ec.europa.eu/system/files/2019-12/ec_rtd_factsheet-open-science_2019.pdf. Consultato il 15 giugno 2023.
- DE GREGORIO, G. (2021), *The Rise of Digital Constitutionalism in the European Union*, in «International Journal of Constitutional Law», 19(1), pp. 41–70

- ENGLER A. (2021), *How Open-Source Software Shapes Ai Policy*, The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET), <https://www.brookings.edu/articles/how-open-source-software-shapes-ai-policy/>. Consultato il 15 giugno 2023.
- FLORIDI, L., COWLS, J., BELTRAMETTI, M., CHATILA, R., CHAZERAND, P., DIGNUM, V., LUETGE, C., MADELIN, R., PAGALLO, U., ROSSI, F. e SCHAFER, B. (2021), *An Ethical Framework for a Good Ai Society: Opportunities, Risks, Principles, and Recommendations*, in «Minds & Machines» 28, 689-707.
- GUNDERSEN, O.E., GIL, Y. e AHA D.W. (2018a), *On Reproducible Ai: Towards Reproducible Research, Open Science, and Digital Scholarship*, in «Ai Publications. Ai Magazine», 39(3), pp. 56-68.
- GUNDERSEN, O.E. e KJENSMO S. (2018b), *State of the Art: Reproducibility in Artificial Intelligence*, in «Proceedings of the AAAI Conference on Artificial Intelligence», doi: org/10.1609/aaai.v32i1.11503.
- HAIBE-KAINS, B., ADAM, G.A., HOSNY, A., KHODAKARAMI, F., MASSIVE ANALYSIS QUALITY CONTROL (MAQC), SHRADDHA, T., KUSKO R., SANSONE, S.A., TONG, W., WOLFINGER, R. D., MASON, C. E., JONES, W., DOPAZO, J., FURLANELLO, C., WALDRON, L., WANG, B., MCINTOSH, C., GOLDENBERG, A., KUNDADJE, A. e GREENE, C.S. (2020), *Transparency and Reproducibility in Artificial Intelligence*, in «Nature», 586(7829), pp. E14-E16.
- HEAVEN, W.D. (2020), *Ai is Wrestling with a Replication Crisis*, in «MIT Technology Review», 12 novembre, <https://www.technologyreview.com/2020/11/12/1011944/artificial-intelligence-replication-crisis-science-big-tech-google-deepmind-facebook-openai/>.
- LINDBERG, V. (2023), *Building and Using Generative Models under Us Copyright Law*, in «Rutgers Business Law Review», 18(2), pp. 1-65.
- NARANG, S. e AAKANKSHA C. (2022), *Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance*, in «Google Research», 4 aprile, <https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>.
- PATEL, D. e AHMAD, A. (2023), *Google «We Have No Moat, And Neither Does OpenAI»*, in SemiAnalysis, 4 maggio, <https://www.semianalysis.com/p/google-we-have-no-moat-and-neither>.
- TEGMARK, M. e OMOHUNDRO, S. (2023), *Provably Safe Systems: The Only Path to Controllable AGI*, in «Arxiv», doi:10.48550/arXiv.2309.01933.