

Milano, febbraio 12, 2024

Posizione INDICAM sulla raccolta di dati personali on line tramite *web scraping* per finalità di addestramento degli algoritmi.

INDICAM – Associazione Italiana per la tutela della proprietà intellettuale (di seguito “INDICAM” o l’“Associazione”)– ringrazia l’Autorità Garante per la protezione dei dati personali (in seguito, anche “Garante Privacy”) per l’opportunità di presentare la propria posizione su questa importante consultazione.

A. Breve presentazione di INDICAM

INDICAM è l’Associazione italiana per la tutela della proprietà intellettuale e la lotta alla contraffazione. Dal 1987 si occupa di supportare le imprese associate nella tutela della loro proprietà intellettuale. Si tratta di imprese operanti in diversi settori merceologici, dalla moda al lusso, all’*automotive*, al *food& beverage*, all’arredamento e *design*, imprese che con il loro patrimonio rappresentano quasi il 2.8% del PIL nazionale.

INDICAM opera come punto di sintesi tra l’industria e tutta una serie di interlocutori nazionali ed internazionali quali: istituzioni, forze dell’ordine, consumatori, piattaforme digitali e grazie al *know-how* sviluppato in questi ormai 35 anni di storia siamo diventati un osservatorio privilegiato su tutti i temi che afferiscono alla proprietà intellettuale

Più specificamente, l’Associazione si propone di identificare i *trend*, le soluzioni, le tecniche più efficaci e innovative per aumentare la consapevolezza e la protezione dei citati diritti.

B. La nostra posizione

1. Finalità di tutela di un diritto in sede di giudizio: liceità del *web scraping* nel contrasto alla pirateria digitale attraverso sistemi di IA

Come noto, il “*web scraping*” è la tecnica informatica che, attraverso l’utilizzo di specifici software, consente l’estrazione massiva di dati e informazioni dai siti web, al fine di utilizzare quanto estratto in altri contesti: a mero titolo esemplificativo, l’arricchimento di *database*, l’indicizzazione delle informazioni o l’addestramento di algoritmi.

A nostro avviso, l’utilizzo della descritta tecnica potrebbe comportare non solo rischi, ma anche benefici a seconda della finalità effettivamente perseguita. Si faccia il caso della finalità di indicizzazione e reperibilità delle informazioni (cd. “*search*”) che, certamente, produce indubbi vantaggi per il titolare delle informazioni con rischi residuali per la privacy degli individui.

Con specifico riferimento alla raccolta massiva di dati per l’addestramento degli algoritmi di intelligenza artificiale, ad avviso di chi scrive, sarebbe opportuno non tanto vietare in assoluto l’utilizzo quanto, piuttosto, regolamentarne le modalità, interrogandosi *in primis* sulla legittimità della pratica *tout court* e individuando, in concreto, i limiti posti a tutela dei diritti riconosciuti dal legislatore per ciò che concerne tanto la protezione dei dati personale quanto la tutela del diritto d’autore.

INDICAM

— PER LA TUTELA DELLA PROPRIETÀ INTELLETTUALE

Affermare che l'implementazione di tecniche di *web scraping* comporti *ex se* una violazione dei diritti avrebbe ricadute dannose, per assurdo, sul piano stesso della tutela del diritto d'autore. La tecnica, infatti, potrebbe essere utilizzata per finalità del tutto legittime, a condizione che vengano bilanciati i diritti in gioco.

È questo il caso della lotta contro la pirateria digitale, fenomeno sempre più diffuso e in parte favorito dall'utilizzo di *web scraping* per le finalità in esame¹, che potrebbe essere efficacemente combattuto utilizzando la stessa arma dell'offensore attraverso misure *anti-piracy* che si basano a loro volta su sistemi di Intelligenza Artificiale (in seguito, per brevità, anche "IA") addestrati al fine di rilevare le violazioni compiute.

Parimenti, è bene rilevare che sistemi di IA basati sulla tecnica di *web scraping* possono consentire una tutela *anti-counterfeiting*, ossia permettono di rilevare se versioni fraudolente di un certo prodotto sono in vendita sul mercato online. L'utilizzo di queste tecnologie può comportare, infatti, una riduzione del rischio di contraffazione e, a differenza di altri strumenti, quali i proxy anti-contraffazione, non è ostacolato da blocchi geografici degli indirizzi IP, dal momento che la rilevazione delle segnalazioni avviene su tutto il web. A maggior ragione, pertanto, la doppia faccia del *web scraping* dev'essere oggetto di accurata valutazione, dovendosi effettuare un *balancing test* di tutti i diritti in gioco.

A ben vedere, il problema è delicato e dev'essere affrontato tenendo bene in considerazione non solo le esigenze di tutela della *privacy* e del diritto d'autore ma anche di diritti equipollenti, quale il diritto di difesa in giudizio.

Sarebbe perciò opportuno avviare un tavolo che coinvolga esperti del settore al fine di definire in quali casi possa ritenersi lecito o meno il *web scraping*.

A tal fine, si coglie l'occasione per chiedere che l'Ill.ma Autorità, confermi la liceità dell'applicazione del *web scraping* per la finalità di identificazione nella rete internet degli illeciti relativi alla contraffazione e più in generale nella violazione del diritto autorale e del diritto industriale.

2. L'approccio "opt-out" della normativa Copyright e il conflitto emerge con l'approccio "opt-in" della normativa sulla data protection.

La Legge sul diritto d'autore (in seguito, anche "L.d.A.") sul diritto d'autore prevede, come noto, agli artt. 96 e 97 che l'immagine di una persona non possa essere esposta, riprodotta o messa in commercio senza il consenso della persona raffigurata².

Tuttavia, la stessa L.d.A. (specificamente all'art. 70-*quater*), recependo la Direttiva (UE) 2019/790 (in seguito, la "Direttiva") prevede un'eccezione per l'utilizzo delle tecniche di "*text and data mining*" laddove è sancito che: "*sono consentite le riproduzioni e le estrazioni da opere o da altri materiali contenuti in reti o in banche di dati cui si ha legittimamente accesso ai fini*

¹ Si richiama, sul punto, lo studio dell'European Union Intellectual Property Office consultabile al seguente URL: https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2023_online_copyright_infringement_in_eu/2023_online_copyright_infringement_in_eu_FullR_en_en.pdf.

² Tranne nel caso – da interpretarsi restrittivamente – in cui vi sia un'esigenza pubblica all'informazione, se giustificata da notorietà o dall'ufficio pubblico coperto, da necessità di giustizia o polizia, da scopi scientifici o didattici e non invece per finalità di lucro, e sempre salvo i casi di lesione a onore, reputazione o decoro.

dell'estrazione di testo e di dati. L'estrazione di testo e di dati è consentita quando l'utilizzo delle opere e degli altri materiali non è stato espressamente riservato dai titolari del diritto d'autore e dei diritti connessi nonché dai titolari delle banche dati”.

Tale norma delinea un approccio “*opt-out*” per cui è il titolare dei diritti che deve esprimere espressamente la riserva e il diniego all’uso delle sue opere e materiali, anche quando sono coinvolti dati personali, come ad esempio l’immagine, a pena di vederli oggetto di indesiderate estrazioni di testo e dati per le più eterogenee finalità, anche commerciali.

Si evidenzia, sul punto, come la disciplina sulla *data protection* di cui al Regolamento (UE) 2016/679 (cd. “GDPR”) presenti un approccio diametralmente opposto: il GDPR prevede, infatti, che il trattamento di dati personali compiuto tramite *web scraping* per finalità di addestramento degli algoritmi, per essere lecito, debba essere fondato sul consenso specifico ed informato dell’interessato (cd. sistema di “*opt-in*”) ai sensi degli artt. 6, par. 1, lett. a) e 9, par. 2, lett. a) del GDPR. Il consenso assume, infatti, un ruolo fondamentale in quanto il trattamento è effettuato con mezzi automatizzati e con possibili finalità di profilazione degli interessati, ai sensi dell’art. 22 del GDPR.

Urge, alla luce di quanto sin qui detto, un intervento dell’Autorità che appiani le antinomie – si spera solo apparenti – tra le normative appena esaminate, considerato che la regolamentazione del fenomeno dovrebbe tendere, quanto più possibile, ad uniformità.

3. Applicazione dell’art. 167-bis, comma 2, per la raccolta illecita di dati personali mediante *web scraping*

Ferme le descritte esigenze di armonizzazione e chiarimento, sembra opportuno domandarsi se l’attuale impianto normativo non consenta già di individuare dei limiti di utilizzo del *web scraping*.

Come noto, l’art. 167-bis del D. Lgs. n. 196/2003 (cd. “Codice Privacy”) prevede che: “*salvo che il fatto costituisca più grave reato, chiunque, al fine trarne profitto per sé o altri ovvero di arrecare danno, comunica o diffonde, senza consenso, un archivio automatizzato o una parte sostanziale di esso contenente dati personali oggetto di trattamento su larga scala, è punito con la reclusione da uno a sei anni, quando il consenso dell’interessato è richiesto per le operazioni di comunicazione e di diffusione*”.

Nel caso di specie, come si è osservato, il *web scraping* utilizzato per finalità di addestramento di modelli informatici, sottostanti all’Intelligenza Artificiale si ritiene che integri un trattamento di dati personali qualificabile come illecito, in quanto privo del consenso dell’interessato che resta ignaro delle finalità perseguite e neppure conoscibili dallo stesso *ex post*; ci si domanda, pertanto, se la disposizione appena riportata comporti la responsabilità penale di chi effettua tale trattamento in assenza del consenso (o di idonea base giuridica ai sensi degli artt. 6 e 9 GDPR) essendo in presenza di un trattamento su larga scala al fine di trarne profitto per sé.

Pertanto, sarebbe auspicabile che l’Autorità nel novero della consultazione *in fieri* chiarisca anche detto aspetto.

4. Misure tecniche e organizzative per la tutela degli interessati e del diritto d'autore

Ai fini della tutela della protezione dei dati personali, si intende valorizzare la possibilità per l'interessato di esprimere un consenso specifico alla finalità di addestramento degli algoritmi mediante *web scraping* in coerenza con l'approccio ("*opt-in based*") del GDPR.

Si propone, inoltre, a fronte dei chiarimenti necessari sul piano della liceità del trattamento in esame in assenza del consenso dell'interessato e sull'applicazione dell'art. 167-bis del Codice Privacy, di prevedere, in ottica di maggiore trasparenza e *accountability*, un obbligo per i fornitori dei predetti *software* di pubblicare le rispettive DPIA (o almeno un estratto delle stesse) e di iscrivere i rispettivi prodotti in un registro pubblico contenente informazioni sulle modalità del trattamento, le tipologie di dati trattati e le misure di sicurezza predisposte dal fornitore.

Ai fini della tutela del diritto d'autore, si consideri l'opportunità dell'adozione delle seguenti misure di sicurezza:

- aggiornamento dei termini di servizio per esplicitare chiaramente in quali casi le attività di *web scraping* sono vietate ed eventualmente perseguibili;
- implementazione di strumenti preventivi per la classificazione e/o analisi del traffico;
- risoluzione dei CAPTCHA/reCAPTCHA che saggiano l'identità "umana" o "automatizzata" dell'utente;
- l'uso di file robots.txt, che indica ai crawler dei motori di ricerca a quali URL del sito possano accedere, specificando i singoli sottodomini "fruibili" o "vietati" per gli user-agent, come nell'esempio riportato di seguito:
- l'inserimento di porzioni di codice (es. html) che vietino la deindicizzazione dei contenuti, come "noindex" e/o "disallow", ovvero un set di regole, con un tag o un'intestazione di risposta http, che viene utilizzato per impedire l'indicizzazione da parte dei motori di ricerca che supportano la regola (es. Google).

Si tenga conto che nell'implementazione di tali misure di sicurezza dovranno esser contemperati e valutati anche i potenziali rischi che possano derivarne, come ad esempio la possibile limitazione nella indicizzazione dei contenuti da parte dei motori di ricerca e/o l'eventuale limitazione dell'attività di antipiracy.

INDICAM, ringraziando per l'opportunità e l'attenzione, rimane a disposizione per ulteriori approfondimenti e per eventuale audizione.

Cordiali saluti

Juna Shehu



Direttore Generale