OSSERVATORIO
TMT · DATA PROTECTION
di Morri Rossetti

Monthly Roundup

Giugno 2024

Giugno 2024

I principali aggiornamenti in materia di TMT & Data Protection del mese.

NUOVI PROVVEDIMENTI LEGISLATIVI E REGOLATORI

Provvedimenti del Garante Privacy

- Ricerca scientifica: le Faq del Garante Privacy per gli IRCCS [Link]
- Garante Privacy: il GDPR vale anche per Wikipedia [Link]
- GDPR e intelligenza artificiale, il Report della task force europea su ChatGPT [Link]
- Telemarketing, dal Garante sanzione di oltre 6 milioni di euro a Eni Plenitude [Link]
- Riconoscimento facciale: Garante sanziona una concessionaria [Link]

EDPB

- Guideline 01/2023 on Article 37 Law Enforcement Directive [Link]

The main challenges for generative AI privacy compliance: the EDPB new report on ChatGPT Taskforce

On April 13, 2023, the European Data Protection Board ("**EDPB**") established a taskforce to foster cooperation and exchange information on possible enforcement actions on the processing of personal data in the context of ChatGPT ("**ChatGPT Taskforce**"). As is commonly known, ChatGPT is the generative artificial intelligence ("**G-AI**") service provided by the company OpenAI OpCo, LLC.

The work of ChatGPT Taskforce has led to the creation of a Report on the work of the ChatGPT taskforce, which was recently adopted by the EDPB ("**Report**"). The Report provides the preliminary views of the ChatGPT Taskforce without prejudging the analysis that the Supervisory Authorities will have to conduct in their respective investigations.

The Report highlights several important privacy related issues that could potentially impact all developers and deployers of G-AI solutions. Indeed, the algorithms of such Large Language Models ("**LLMs**") are trained using the so-called web scraping technique, which enables the automated collection and extraction of various types of information (including, personal data and even special categories of personal data) from different publicly available sources on the Internet.

Additionally, the taskforce members have developed a **common questionnaire** as a possible basis for their exchanges with Open AI, which is published as an annex to the Report. This set of questions aims to promote a coordinated approach to the investigation and can also serve as a useful guide for other providers in developing G-AI systems compliant with the data protection regulation.

But which are the principles on personal data applicable to ChatGPT?

Lawfulness

Generally speaking, the EDPB recalls that each processing of personal data must meet one of the legal basis set forth in Article 6(1) of the Regulation (EU) 2016/679 ("**GDPR**") and, where applicable, the additional requirements for processing special categories of personal data pursuant to Article 9(2) of the GDPR.

The Report details the use of **legitimate interest** for the collection and processing of personal data to train ChatGPT and sets out the **limits** within which, according to the EDPB, this could be considered acceptable. Indeed, the EDPB emphasises that while **legitimate interest** may potentially be used as legal basis, it should be based on a proper legitimate **interest assessment (LIA)**, and **adequate safeguards** to reduce undue impact on data subjects should be implemented, potentially changing the balancing test in favour of the controller. Such safeguards could include, for example:

a) **technical measures**, defining precise collection criteria;
b) ensuring some categories of personal data are not collected or some sources (e.g. public social media profiles) **are excluded from data collection;**

c) **erasure and anonymisation** of personal data collected via web scraping before the training stage.

Fairness

The principle of fairness requires that personal data should not be processed in a manner that is unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject.

A crucial aspect of this principle is that there should be no risk transfer. This means that **ensuring compliance with the GDPR is a responsibility of OpenAI and not of the data subjects, even when individuals input personal data**.

Information obligations, transparency and data accuracy

The Report outlines that if personal data are collected via web scraping from publicly accessible sources or while directly interacting with ChatGPT, the controller should provide **proper information** to the data subjects.

Given the vast amount of data collected via web scraping, it is often not practicable or possible to inform each data subject individually. Thus, the controller shall provide all the information set forth in art. 14(1) and (2) of the GDPR, such as general information and contact details of the controller, categories of personal data being processed, data retention period, rights of the data subjects, etc.

Conversely, when personal data are collected directly from the data subject (i.e. pursuant to art. 13 of the GDPR), the controller shall inform them that the content (i.e. the prompt, the uploaded files and ChatGPT responses) is used to train and improve the LLM.

Furthermore, due to the probabilistic nature of ChatGPT, the EDPB highlights that the controller should not only provide proper information about the probabilistic output creation mechanism and its limited reliability, but also disclose that the generated text, although syntactically correct, may be biased or made up.

Rights of the data subjects

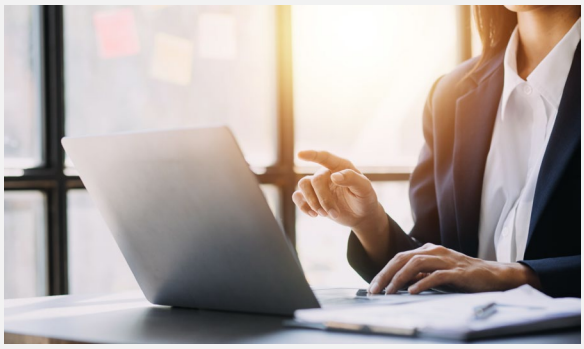
The Report stresses the **importance of data subjects being able to effectively exercise their rights.** While OpenAI, as controller, provides information on how to exercise these rights in its privacy policy[1], the EDPB – pursuant to art. 12(2) and Recital 59 of the GDPR – asserts that the controller shall continue to improve the modalities for facilitating the exercises of such rights. This is particularly relevant as OpenAI suggests shifting from rectification to erasure when rectification is not feasible due to the technical complexity of ChatGPT.

In line with the principles of privacy by design and by default, the controller shall adopt appropriate measures, both when determining the means of processing and when processing itself, to effectively implement data protection principles and integrate the necessary safeguards into processing to meet GDPR requirements and protect the rights of data subjects.

* * *

[1] Europe privacy policy of OpenAI, paragraph 6.

DP & Labour Alert | Posta elettronica e metadati: il nuovo documento di indirizzo del Garante Privacy [2]



L'Autorità Garante per la protezione dei dati personali (il "**Garante Privacy**") ha pubblicato un nuovo documento di indirizzo, adottato lo scorso 6 giugno, denominato "Programmi e servizi informatici di gestione della posta elettronica nel contesto lavorativo e trattamento dei metadati" (il "**Documento**").

Il Garante Privacy, a seguito delle osservazioni e delle proposte pervenutegli nell'ambito della consultazione pubblica indetta con provvedimento del 22 febbraio scorso, ha così aggiornato il precedente documento in materia (per maggiori approfondimenti, si veda il nostro precedente contributo disponibile qui), adottato in via preliminare allo scopo di richiamare l'attenzione su alcuni punti di intersezione tra la normativa in materia di protezione dei dati personali (i.e. il Regolamento (UE) 679/2016 – il "**GDPR**" e il D.Lgs. 196/2003, come modificato dal D.Lgs. 101/2018 – il "**Codice Privacy**") e le norme che stabiliscono le condizioni per l'impiego degli strumenti tecnologici nei luoghi di lavoro (i.e. la L. 300/1970 e ss.mm.ii. – lo "**Statuto dei Lavoratori**").

L'obiettivo sotteso ai due documenti è, dunque, il medesimo: fornire indicazioni in merito al rischio, insito in alcuni programmi e servizi di gestione della posta elettronica (anche qualora commercializzati in modalità cloud) e relativo all'utilizzo di tali account in uso ai dipendenti, di raccolta dei metadati per impostazione predefinita, in modo preventivo e generalizzato e conservando gli stessi per un arco temporale troppo esteso.

Il Documento ha **natura** meramente **orientativa** e non impone, dunque, in capo ai titolari del trattamento, alcun nuovo adempimento o responsabilità.

Quali sono le principali novità del Documento?

1. **Una definizione di "metadati" più chiara e precisa**

Il Garante Privacy precisa che i metadati a cui fa riferimento il Documento corrispondono alle **informazioni registrate nei log generati dai server di gestione e smistamento della posta elettronica (*Mail Transport Agent*) e dalle postazioni nell'interazione che avviene tra i diversi server interagenti e, se del caso, tra questi e i *client*** (i.e. le postazioni terminali che effettuano l'invio dei messaggi e che consentono la consultazione pubblica della corrispondenza in entrata accedendo a mailbox elettroniche – Mail *User Agent*).

Tali informazioni possono comprendere **gli indirizzi e-mail del mittente e del destinatario, gli indirizzi IP dei server o dei *client*** coinvolti nell'instradamento dei messaggi, gli **orari di invio, di ritrasmissione o di ricezione, la dimensione del messaggio, la presenza e la dimensione di eventuali allegati** e, in certi casi, in base al sistema di gestione del servizio di posta elettronica utilizzato, anche **l'oggetto del messaggio spedito o ricevuto.**

Si tratta, dunque, di metadati che condividono tutti la caratteristica di essere registrati

[2] Il presente contributo è stato redatto in collaborazione con il Team del Dipartimento di Diritto del Lavoro e delle Relazioni Industriali

automaticamente dai sistemi di posta elettronica, indipendentemente dalla percezione e volontà dell'utente.

In tale ottica, i metadati oggetto del Documento **non devono in alcun caso essere confusi con quelle informazioni che costituiscono il corpo del messaggio o integrate nello stesso** in modo da formare il c.d. envelope, ossia l'insieme delle intestazioni tecniche strutturate che documentano l'instradamento del messaggio, la sua provenienza e altri parametri tecnici e che restano inscindibili dal messaggio stesso.

2. **La necessità di rispettare la normativa in materia di protezione dei dati personali**

Il Garante Privacy richiama l'applicabilità dei seguenti principi:

- **principio di riservatezza**, in quanto il contenuto dei messaggi di posta elettronica, i dati esteriori delle comunicazioni e i file allegati, riguardano forme di corrispondenza assistite da garanzie di segretezza tutelate anche dalla Costituzione (artt. 2 e 15 Cost.);

- **principio di liceità**: il datore di lavoro, in qualità di titolare del trattamento, deve verificare la sussistenza dei presupposti di liceità posti dall'art. 4 dello Statuto dei Lavoratori[1] prima di effettuare trattamenti di dati personali dei lavoratori attraverso programmi e servizi di gestione della posta elettronica;

- **principio di trasparenza**, in base al quale il titolare del trattamento – al fine di fornire una rappresentazione chiara del trattamento effettuato – deve rispettare i principi generali del trattamento e attuare tutti gli adempimenti legali di trasparenza previsti dal GDPR (ad esempio, gli obblighi informativi di cui all'art. 13 del GDPR);

- **principio di accountability**, in virtù del quale il titolare del trattamento deve valutare se i trattamenti che intende realizzare possano presentare un rischio elevato per i diritti e le libertà delle persone fisiche, rendendosi così necessaria una preventiva valutazione d'impatto sulla protezione dei dati personali ("**DPIA**").

3. **La necessità di rispettare le finalità e le garanzie poste dall'art. 4 dello Statuto dei Lavoratori e la previsione di un nuovo periodo di conservazione dei metadati**

L'art. 4, comma 2, dello Statuto dei Lavoratori prevede espressamente che gli strumenti preordinati allo svolgimento della prestazione lavorativa, in quanto funzionali all'esecuzione della stessa, non soggiacciono al regime normativo previsto dal primo comma della medesima disposizione – il quale richiede, invece, che gli altri impianti e strumenti dai quali derivi anche solo la possibilità di controllo a distanza dell'attività dei lavoratori possano essere impiegati dal datore di lavoro solo (i) per specifiche finalità (organizzative, produttive, di sicurezza del lavoro e di tutela del patrimonio aziendale), e (ii) stabilendo precise garanzie procedurali (accordo sindacale o autorizzazione pubblica).

Ebbene, il Garante Privacy afferma che possa ritenersi applicabile l'eccezione di cui al predetto secondo comma dell'art. 4 all'attività di raccolta e conservazione dei soli metadati/log necessari ad assicurare il funzionamento delle infrastrutture del sistema di posta elettronica se la stessa viene effettuata per un periodo limitato, che – a titolo orientativo – **non dovrebbe comunque**

superare i 21 giorni [3] **, salvo specifiche necessità**.

Entro tali limiti di continenza, il Garante reputa che la conservazione di tali dati sia **finalizzata ad assicurare il funzionamento dello strumento funzionale alla prestazione lavorativa** e, quindi, tale trattamento sarebbe effettuabile dal datore di lavoro "semplicemente" dando al lavoratore adeguata informazione delle modalità d'uso di tale strumento e dell'effettuazione dei relativi controlli. La raccolta generalizzata e periodi di conservazione più estesi, invece, potendo comportare un indiretto controllo a distanza dei lavoratori, possono essere adottate soltanto nei limiti e con le garanzie di cui all'art. 4, comma 1 dello Statuto dei Lavoratori.

Da ultimo, vale la pena rammentare che le predette indicazioni del Garante si applichino ai soli "metadati" come definiti dal medesimo documento di indirizzo (ossia, le informazioni relative alle operazioni di invio e ricezione e smistamento dei messaggi), non i **contenuti dei messaggi di posta** che rimangono nella disponibilità del lavoratore all'interno della casella di posta elettronica attribuitagli e per il trattamento dei quali non potrà prescindersi dal rispetto delle **previsioni (più gravose) del primo comma dell'art. 4 citato**.

4. **Le possibili responsabilità dei datori di lavoro**

I datori di lavoro dovranno necessariamente verificare i **presupposti di liceità** del trattamento, garantire la **correttezza** e la **trasparenza** e rispettare il **principio di limitazione della conservazione**, i principi di ***privacy by design*** e ***privacy by default***, nonché il principio di responsabilizzazione (***accountability***).

Pertanto, al fine di evitare possibili conseguenze sia sul piano amministrativo che penale, i datori di lavoro dovranno verificare che i programmi e servizi informatici di gestione della posta elettronica gli consentano di rispettare tali principi, anche con riguardo al periodo di conservazione dei metadati previsto dal Documento.

5. **Il ruolo dei fornitori dei servizi di posta elettronica**

Infine, il Garante Privacy pone l'accento sul ruolo dei fornitori dei servizi di posta elettronica. È infatti compito di questi ultimi contribuire a far sì che i titolari del trattamento siano in grado di adempiere ai loro obblighi in materia di protezione dei dati personali.

In altri termini, il Garante Privacy impone anche ai fornitori l'obbligo di **rispettare i principi di *privacy by design* e *privacy by default***. Ciò significa, dunque, che già in fase di progettazione, sviluppo, selezione e utilizzo di applicazioni, servizi e prodotti basati sul trattamento dei dati personali o che trattano dati personali, i produttori di tali servizi e/o applicazioni dovranno contemperare le proprie esigenze di commercializzazione su larga scala con la conformità di tali prodotti ai principi del GDPR.

* * *

[3] Nel precedente documento, il Garante Privacy aveva invece previsto un periodo di conservazione non superiore a 7 giorni.

Web scraping e IA generativa: come prevenire la raccolta non autorizzata di dati personali?

Nel panorama tecnologico odierno, l'intelligenza artificiale generativa ("**IAG**") rappresenta sicuramente una delle tecnologie più promettenti, con benefici innegabili in termini di efficienza, celerità e perfezionamento della qualità del lavoro. Tuttavia, essa è al contempo una delle tecnologie più contestate e ciò per le sue implicazioni – talvolta negative – in materia di protezione dei dati personali.

Gli algoritmi di IAG richiedono infatti notevoli quantità di dati (anche di carattere personale) per il loro addestramento, spesso provenienti da una raccolta massiva ed indiscriminata effettuata sul *web*. Tra le diverse pratiche utilizzate per addestrare tali sistemi, una delle più diffuse è quella del *web scraping*.

In tale scenario, l'Autorità garante per la protezione dei dati personali (il "**Garante Privacy**") è intervenuta indicendo un'indagine conoscitiva in materia di *web scraping*, la quale ha condotto all'emanazione, lo scorso 20 maggio, di una specifica nota informativa sul web scraping e l'IAG (la "**Nota Informativa**").

L'obiettivo del Garante Privacy è quello di segnalare possibili azioni di contrasto che i gestori di siti internet e di piattaforme online, sia pubblici che privati, operanti in Italia, quali titolari del trattamento di dati personali oggetto di pubblicazione, potrebbero implementare al fine di prevenire – ove ritenuta incompatibile con le basi giuridiche e le finalità della pubblicazione – la raccolta di dati da parte di terzi per finalità di addestramento dei modelli di IAG.

Ma cos'è il *web scraping* e quali problematiche in materia di protezione dei dati personali solleva?

Il *web scraping* è una particolare tecnica utilizzata per raccogliere, memorizzare e conservare, in maniera sistematica e automatizzata, una quantità massiva e indiscriminata di informazioni e dati pubblicamente disponibili *online* o resi disponibili in aree ad accesso controllato. I dati raccolti vengono successivamente impiegati per mirate analisi, elaborazioni e utilizzi.

Con l'avvento dell'IAG, la tecnica informatica del *web scraping* ha avuto una crescita esponenziale, consentendo una raccolta automatizzata di dati più veloce e trasversale, per poi utilizzarli per l'addestramento dell'IAG stessa. Le informazioni che tali tecniche sono in grado di estrarre sono molteplici e tra queste sicuramente rientrano anche dati personali. Si pensi, ad esempio, ai dati di contatto, ai dati biometrici e di geolocalizzazione, alle preferenze personali o anche ai comportamenti di navigazione. In tali ipotesi, i.e. quando il *web scraping* implica la raccolta di informazioni riconducibile a una persona identificata o identificabile, si pone un problema di protezione dei dati personali (per un maggior approfondimento sugli ulteriori profili legali connessi e derivanti dalle attività di web scraping, si rinvia ad un nostro precedente contributo, disponibile qui).

Nello specifico, in questi casi, il focus della compliance si incentra sulla necessità di individuare un'idonea base giuridica per il trattamento di tali dati e sul rispetto dei principi generali posti dal Regolamento (UE) 679/2016 (il "**GDPR**"). Ciò significa dunque che i gestori di siti *web* e di piattaforme online che rivestano al

contempo il ruolo di titolare del trattamento dovranno rispettare gli obblighi di trasparenza, pubblicità, riutilizzo, accesso e adozione delle necessarie misure di sicurezza. Infatti, il fatto che i dati personali siano pubblicamente reperibili, non equivale ad acconsentire ad un loro libero utilizzo.

Le indicazioni del Garante Privacy ai gestori di siti *web* e piattaforme *online*

Al netto degli ulteriori obblighi gravanti sui titolari del trattamento posti dal GDPR, il Garante Privacy, con la propria Nota Informativa, ha voluto fornire alcune indicazioni ai gestori dei siti *web* e di piattaforme online in merito alle possibili cautele che gli stessi potrebbero adottare per mitigare gli effetti del *web scraping* di terze parti finalizzato all'addestramento di sistemi di IAG.

Nello specifico, il Garante Privacy ha individuato quattro diverse misure (contenitive, ma non risolutive), aventi carattere tecnico, tecnico-organizzativo e legale:

1. la **creazione di aree riservate**, ossia la predisposizione di aree del sito o delle piattaforme a cui è possibile accedere solo previa registrazione, sottraendo così i dati dalla pubblica disponibilità. Il Garante Privacy sottolinea però come, di contro, tale misura non possa dar luogo ad un trattamento di dati eccessivo da parte del titolare, in violazione del principio di minimizzazione di cui all'art. 5(1)(c) del GDPR, ad esempio imponendo agli utenti oneri di registrazione ulteriori e ingiustificati;

2. **l'inserimento di clausole *ad hoc* nei termini di servizio**, adottando una cautela di mera natura giuridica, operante a posteriori. Infatti, nel caso in cui tali clausole non venissero rispettate, i gestori dei siti e delle piattaforme sarebbero legittimati ad agire in giudizio per far dichiarare l'inadempimento contrattuale della controparte;

3. il **monitoraggio del traffico di rete**, mediante un accorgimento tecnico che è in grado di individuare eventuali flussi anomali di dati in ingresso e in uscita da un sito *web* o da una piattaforma *online* e, conseguentemente, di adottare adeguate contromisure di protezione;

4. **l'intervento sui bot.** Posto che il web scraping si basa sull'utilizzo dei bot, il Garante Privacy sottolinea come una qualsiasi tecnica in grado di limitare l'accesso agli stessi rappresenti un metodo efficace per arginare l'attività automatizzata di raccolta dati effettuata tramite tali *software*.

Tra tali tecniche rientrano, a mero titolo esemplificativo, l'inserimento di verifiche CAPTCHA (*Completely Automated Public Turing-test-to-tell Computers and Humans Apart*), che impongono un'azione eseguibile solo da un essere umano, la modifica periodica del markup HTML, oppure l'incorporazione dei contenuti o dei dati all'interno di oggetti multimediali, come ad esempio, le immagini.

Tuttavia, come evidenziato dal Garante Privacy, nessuna di tali misure può essere considerata sufficiente e idonea ad impedire completamente le tecniche di *web scraping*. Si tratta, dunque, di cautele che devono essere adottate sulla base di un'autonoma valutazione del titolare del trattamento, da svolgersi caso per caso in base allo specifico contesto di riferimento e in attuazione del principio di *accountability*, nonché nel rispetto dei principi in materia di protezione dei dati personali previsti dal GDPR.

L'intervento del Garante Privacy è cruciale per favorire una maggior consapevolezza nell'utilizzo degli strumenti di IAG da parte delle aziende. In tale ottica, sarà fondamentale garantire un approccio bilanciato e multidisciplinare.

Se infatti, da un lato, le diverse misure tecniche e legali potrebbero fungere da deterrente per le pratiche di *web scraping* non autorizzate, dall'altro ciò potrebbe comportare un rallentamento dell'innovazione delle nuove tecnologie di IAG. Pertanto, nell'adozione delle stesse, occorre porre molta cautela, bilanciando gli interessi in gioco e adottando misure di prevenzione e mitigazione che siano proporzionate e non eccessivamente onerose, dovendo coinvolgere in tale processo tutti i soggetti, tecnici e legali, esperti in materia.

Per maggiori informazioni e approfondimenti

Carlo Impalà
Partner e Responsabile Osservatorio TMT&DP

Carlo.Impala@MorriRossetti.it

Morri Rossetti

Osservatorio TMT&DP

OSSERVATORIO
TMT · DATA PROTECTION
di Morri Rossetti

Piazza Eleonora Duse, 2
20122 Milano
MorriRossetti.it

Osservatorio-dataprotection.it