



**GARANTE  
PER LA PROTEZIONE  
DEI DATI PERSONALI**

## **Provvedimento del 20 maggio 2024 [10020316]**

**VEDI ANCHE:** [Comunicato stampa del 30 maggio 2024](#)



[Web scraping ed intelligenza artificiale generativa - Nota informativa](#)

[doc. web n. 10020316]

### **Provvedimento del 20 maggio 2024**

*(In corso di pubblicazione sulla Gazzetta Ufficiale)*

Registro dei provvedimenti  
n. 329 del 20 maggio 2024

### **IL GARANTE PER LA PROTEZIONE DEI DATI PERSONALI**

NELLA riunione odierna, alla quale hanno preso parte il prof. Pasquale Stanzone, presidente, la prof.ssa Ginevra Cerrina Feroni, vicepresidente, il dott. Agostino Ghiglia e l'avv. Guido Scorza, componenti, e il cons. Fabio Mattei, segretario generale;

VISTO il 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 (Regolamento generale sulla protezione dei dati, di seguito, "Regolamento");

VISTO il Codice in materia di protezione dei dati personali, recante disposizioni per l'adeguamento dell'ordinamento nazionale al Regolamento (UE) 2016/679 (d.lgs. 30 giugno 2003, n. 196, come modificato dal d.lgs. 10 agosto 2018, n. 101, di seguito "Codice");

VISTO l'articolo 57, par. 1, lett. b), del Regolamento che attribuisce alle autorità di controllo il compito di promuovere la consapevolezza e favorire la comprensione del pubblico riguardo ai rischi, alle norme, alle garanzie e ai diritti in materia di protezione dei dati personali, con particolare attenzione alle attività destinate specificamente ai minori;

RILEVATO che è nota l'attività di raccolta massiva, attraverso tecniche di web scraping, da parte di soggetti che sviluppano sistemi di intelligenza artificiale generativa (di seguito anche "IAG") di grandi quantità di dati, anche personali, pubblicati nei siti web e nelle piattaforme online gestiti da soggetti pubblici e privati, stabiliti nel territorio italiano, per finalità di addestramento di tale tipologia intelligenza artificiale;

CONSIDERATO che restano fermi in capo ai soggetti che trattano dati personali, direttamente od indirettamente raccolti tramite tecniche di web scraping per finalità di addestramento di IAG, nella misura in cui essi siano qualificabili come titolari del trattamento, tutti gli obblighi previsti dal Regolamento, in particolare l'obbligo di individuare una idonea base giuridica per il trattamento di

tali dati, in ossequio al principio di accountability di cui all'art. 5, par. 2, del Regolamento;

CONSIDERATO che restano ferme le disposizioni in materia di obblighi di pubblicazione per finalità di trasparenza di cui al d. lgs. n. 33/2013 e altre pubblicità legali, in materia di apertura dei dati e riutilizzo dell'informazione del settore pubblico ai sensi d. lgs. n. 36/2006 (s.m.i.), in materia di prevenzione della corruzione e trasparenza da parte di società ed enti di diritto privato controllati e partecipati dalle pubbliche amministrazioni e di enti pubblici economici, nonché le disposizioni previste da normative specifiche come quelle a tutela della proprietà intellettuale e del diritto d'autore;

CONSIDERATO che restano fermi altresì, in quanto applicabili, i regolamenti (cosiddetti Digital Markets Act, Digitale Service Act, Digital Governance Act e Data Act) che hanno dato attuazione alla strategia digitale europea finalizzata alla creazione di uno spazio comune europeo dei dati in grado di stimolare l'innovazione digitale nell'ambito del mercato unico europeo;

CONSIDERATO che grava sui gestori, pubblici e privati, di siti web e piattaforme online, stabiliti in Italia, l'onere di rispettare i principi fondamentali previsti dal Regolamento e, in particolare, di quelli di limitazione della finalità, di minimizzazione dei dati e di integrità e riservatezza, ai sensi dell'art. 5, par. 1, lett. b), c) ed f) del Regolamento;

TENUTO CONTO delle peculiarità dell'innovazione tecnologica sottesa ai sistemi di IAG nella misura in cui essa comporta dei rischi per la protezione dei dati personali, in particolare con riferimento alle difficoltà in capo agli interessati di esercitare in maniera efficace i diritti di cui agli artt. da 13 a 22 del Regolamento;

TENUTO CONTO dei contributi ricevuti dall'Autorità nell'ambito dell'[indagine conoscitiva](#) in materia di web scraping, deliberata con provvedimento del 21 dicembre 2023, pubblicato nella Gazzetta Ufficiale n. 14 del 18 gennaio 2024;

CONSIDERATO che taluni soggetti che sviluppano sistemi di intelligenza artificiale generativa hanno essi stessi messo a disposizione dei gestori, pubblici e privati, di siti web e piattaforme online misure tecnologiche che consentono di escludere, in tutto o in parte, il contenuto dei loro siti e piattaforme dall'attività web scraping indesiderato;

CONSIDERATO che la pubblicazione di dati personali online da parte di qualunque titolare pubblico o privato di dati personali deve essere sempre basata su una o più basi giuridiche tra quelle indicate all'art. 6 del Regolamento e deve sempre avvenire per una o più specifiche finalità normalmente diverse dalla costituzione di basi di dati aperte utilizzabili da terzi per l'addestramento di algoritmi di intelligenza artificiale;

RITENUTO utile fornire prime indicazioni sulle tecniche di raccolta massiva di dati personali dal web per finalità di addestramento dei modelli di intelligenza artificiale generativa ed indicare ai gestori dei siti web e di piattaforme online, stabiliti in Italia, nella misura in cui siano qualificabili come titolari del trattamento, possibili azioni di contrasto che potrebbero essere adottate per mitigare gli effetti del web scraping di terze parti finalizzato all'addestramento di sistemi di intelligenza artificiale generativa che venisse ritenuto incompatibile, dal singolo titolare del trattamento, in attuazione del principio di accountability, con le finalità e la base giuridica della pubblicazione dei dati personali nonché con eventuali disposizioni di legge speciali vigenti di terze parti finalizzato all'addestramento di sistemi di intelligenza artificiale generativa;

RILEVATO che dette misure non debbono essere intese come obblighi di messa in conformità al Regolamento bensì come misure tecniche ed organizzative la cui adozione dovrà essere valutata tenuto conto anche dello stato dell'arte e dei costi di attuazione (in particolare con riferimento alle PMI), nonché della natura, dell'ambito di applicazione, del contesto e delle finalità dei trattamenti

effettuati e tenuto altresì conto che si tratta di misure non esaustive, da un punto di vista tecnologico, rispetto al fenomeno del web scraping indesiderato;

VISTE le osservazioni formulate dal segretario generale ai sensi dell'art. 15 del regolamento n.1/2000;

RELATORE il prof. Pasquale Stanzione;

### **DELIBERA**

1) ai sensi dell'art. 57, par. 1, lett. b), del Regolamento, di adottare la "[Nota informativa](#)" in materia di web scraping per finalità di addestramento di intelligenza artificiale generativa e di possibili azioni di contrasto a tutela dei dati personali" contenuta nel documento allegato che forma parte integrante della presente deliberazione.

2) ai sensi dell'art. 154 bis, comma 3, del Codice, che copia del presente provvedimento sia trasmessa al Ministero della giustizia per la sua pubblicazione sulla Gazzetta Ufficiale della Repubblica italiana, a cura dell'Ufficio competente.

*Roma, 20 maggio 2024*

IL PRESIDENTE  
Stanzione

IL RELATORE  
Stanzione

IL SEGRETARIO GENERALE  
Mattei