Differential Privacy: la nuova frontiera per il miglioramento della privacy

di Roberta Nobile

Abstract: *Differential Privacy: the new frontier in privacy enhancement* - In order to cope with the limitations of the data de-anonymisation process, Differential Privacy techniques allow for the anonymisation of data already in the acquisition phase, including a set of procedures designed to guarantee the security of any personal data contained in the datasets used for the training of artificial intelligence systems. Lastly, an analysis of the potential and limitations of the GDPR will be conducted, observing the US and New Zealand's application approach in a comparative key, within a comparison in which the New Zealand Privacy Act 2020 strikes a balance between the US modus operandi and the GDPR.

Keywords: Differential privacy; Anonymisation; New Zealand; USA; GDPR

1. Quale privacy per quale futuro?

In una società sempre più digitalizzata e data driven come quella odierna, argomenti come big data, intelligenza artificiale, algoritmi e machine learning spesso aprono profili di incertezza circa la sicurezza delle informazioni e la privacy dei dati personali trattati. Infatti, una parte sempre più consistente della nostra vita finisce online. Molti di questi dati vengono utilizzati per addestrare le intelligenze artificiali a svolgere compiti sempre più complessi, con la massima accuratezza possibile. Una delle sfide dei nuovi algoritmi di Machine learning consiste nell'esigenza di garantire la privacy e la sicurezza dei dati utilizzati nella fase di training. L'obiettivo finale è, infatti, quello di consentire alle intelligenze artificiali di apprendere dei dati senza includere nel modello informazioni sensibili. L'approccio della Differential Privacy[1] al problema può rappresentare una soluzione rivoluzionaria, fornendo, altresì, uno strumento in grado di valutare le garanzie di privacy offerte dal sistema, nel caso di un composition attack[2].

[1] L. Ninghui, et al., *Differential privacy: From theory to practice*, New York, 2017; C. Dwork, A. Roth, *I fondamenti algoritmici della privacy differenziale*, in *Fondamenti e tendenze nell'informatica teorica* 9.3–4, 2014; M. Abadi, et al., *Deep learning with differential privacy*, in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016; P.K. Das, H.K. Tripathy, S.A.M. Yusof, *Privacy and Security Issues in Big Data*, Singapore, 2021.

[2] D.J. Solove, P.M. Schwartz, *Information privacy law*, Boston, 2020; R. Wacks, *Privacy: A very short introduction*, Oxford, 2015; J. Medková, *Composition attack against social network data*, in *Computers & Security* 74, 2018.

DPCE online
ISSN: 2037-6677

Le tecniche di Differential Privacy comprendono, infatti, un set di procedure pensate per garantire la sicurezza e la riservatezza degli eventuali dati personali contenuti nei dataset utilizzati per il training dei sistemi di intelligenza artificiale[3]. L'idea di base è quella di aggiungere "rumore" ai dati, con l'introduzione di un certo grado di entropia nel dataset, tale per cui le informazioni contenute non siano più identificabili o correlabili. In questo modo, si renderebbe inapplicabile il reverse engineering, ossia la tecnica che, partendo da informazioni anonime correlate ad altre che non lo sono, riesce a risalire al dato completo e all'identità dell'utente. Infatti, il punto di forza della Differential Privacy risiede proprio in quella specifica tecnologia che permette di ricevere il dato già anonimizzato, il che mette al riparo da qualsiasi violazione e rende il soggetto che raccoglie e tratta tale dato perfettamente compliant al GDPR.

Il presente paper si propone di sviluppare una disamina sulla nuova prospettiva aperta dalla Differential Privacy sul profilo della tutela della privacy, esaminando, in primo luogo, le relative potenzialità e gli aspetti di criticità e descrivendo, in secondo luogo, alcuni esempi di applicazione, nel tentativo di stabilire se tale "privacy diversa" possa rivelarsi funzionale ad un impiego su larga scala. Singolare risulta essere, come vedremo, l'impiego della di tale tecnica da parte della Nuova Zelanda e degli USA rispettivamente per la protezione dei dati sanitari e per la tutela dei dati ottenuti dal censimento decennale, all'interno di diverse esigenze di ordine medico e politico che contribuiscono ad evidenziare un diverso panorama legislativo: la linearità rivoluzionaria del Privacy Act neozelandese e il mosaico multiforme statunitense, in cui autoregolamentazione e settorialità sfidano il disegno di una futura legislazione sulla privacy.

2. Modelli differenziati di privacy o una privacy diversa?

Come accennato in precedenza, al termine del processo di iniezione del rumore, l'informazione arriva al titolare del trattamento già anonimizzata. Questo è importante in quanto se il dato fosse anonimizzato in un secondo momento, vi sarebbe, tra l'altro, il rischio che il titolare, per svariati motivi, decida di non provvedere all'anonimizzazione; non solo, in secondo luogo, vi sarebbe il pericolo di subire un data breach nel percorso/lasso di tempo che va dalla raccolta del dato alla effettiva anonimizzazione.

Il processo di Differential Privacy risolve entrambe le problematiche rendendo conseguentemente più sicuro il trattamento. È molto significativa questa scelta, specialmente se letta in combinato con la normativa europea sulla data protection la quale, come noto, guarda alla minimizzazione come ad un principio fondamentale imprescindibile[4]. In tal senso, è condivisibile la

[3] R. Marmo, *Algoritmi per l'intelligenza artificiale: Progettazione dell'algoritmo-Dati e Machine Learning-Neural Network-Deep Learning*, Milano, 2020; G.P. Italiano, *Le sfide interdisciplinari dell'intelligenza artificiale*, in *Analisi Giuridica dell'Economia* 18.1, 2019; G. Alpa, *L'intelligenza artificiale. Il contesto giuridico*, Modena, 2021.

[4] R. Cummings, D. Desai, *The role of differential privacy in Gdpr compliance*, in *FAT'18: Proceedings of the Conference on Fairness, Accountability, and Transparency*. 2018; J. Holzel, *Differential Privacy and the GDPR*, in *Eur. Data Prot. L. Rev.* 5, 2019; D.L. Oberski, F.

volontà di anonimizzare il dato anziché limitarsi ad una pseudonimizzazione in quanto, come noto, mentre i dati pseudonimizzati debbono considerarsi comunque dei dati personali (vedi considerando n.26 del GDPR), i dati anonimizzati possono, invece, considerarsi dati non personali in quanto non riconducibili (nemmeno attraverso operazioni di reverse engineering) ad un soggetto determinato.

La Differential Privacy costituisce, quindi, una prospettiva di miglioramento e di tutela delle privacy più efficace rispetto ad altre tecniche conosciute. Volendo partire dal caso della pseudonimizzazione, tale metodo consiste nella sostituzione di dati identificabili con pseudonimi, come ad esempio un ID generato casualmente e collegato ai dati, in modo tale da garantire l'impossibilità di risalire all'identità dell'interessato. Questo offre, sicuramente, alcuni vantaggi in termini di privacy, soprattutto contro forme di divulgazione accidentale, tuttavia, di solito è ancora possibile identificare nuovamente le persone combinandole con informazioni esterne[5], con una relativa elevata e pericolosa probabilità di divulgazione di informazioni sensibili[6]. Infatti, dalla lettura dell'art. 4 comma 5 del GDPR si ricava come la pseudonimizzazione sia una operazione di de-identificazione reversibile, per cui non può costituire una misura di sicurezza a sé stante, ma deve essere inserita entro il complesso delle misure adottate per garantire la sicurezza dei dati[7]. L'anonimizzazione, invece, basandosi sulla rimozione di elementi che permettano di risalire alla persona fisica specifica, rende di norma quasi impossibile la reversibilità del dato[8]. Tuttavia, tale misura può essere realizzata secondo differenti tecniche, due su tutte l'aggregazione e la de-identificazione. Nel primo caso, trattandosi di dati aggregati e, dunque, di una sommatoria di dati di molti individui, la possibilità di una re-identificazione è decisamente remota. Nella seconda, invece, i dati personali sono mantenuti intatti, ma specifiche informazioni di identificazione vengono sostituite con identificatori anonimi. Tale pratica presenta, di conseguenza, dei profili di rischio in termini di identificabilità dell'interessato. Si pensi, per ipotesi, alla banca dati di una prigione che conservi i precedenti penali di un detenuto unitamente alla sua storia medica. Il detenuto, mediante i dati relativi alla fedina penale, potrebbe essere identificato anche senza il nome, e correlativamente si potrebbe facilmente

Kreuter, *Differential privacy and social science: an urgent puzzle*, in *Harvard Data Science Review* 2.1, 2020, 10.

[5] G. D'Acquisto, M. Naldi, *Big data e privacy by design*, Torino, 2017; S. Martinelli, *Il nuovo Regolamento generale sulla protezione dei dati: alcune considerazioni informatico-giuridiche*, in *Diritto mercato tecnologia*, 2016, 1-23.

[6] Si sottolinea come il legislatore europeo abbia optato, al contrario della pseudonimizzazione, per non definire esplicitamente il concetto di anonimizzazione. Nel Considerando 75 del GDPR si parla di rischio di decifratura non autorizzata della pseudonimizzazione. Il legislatore qui indica il fatto che il dato anonimizzato non possa essere ricostruito, mentre quello pseudonomizzato corre il rischio di essere ricostruito qualora fosse possibile riaccoppiare le informazioni.

[7] V. il Considerando 26 del GDPR.

[8] M.V. De Azevedo Cunha, D. Doneda, N. Andrade, *La re-identificazione dei dati anonimi e il trattamento dei dati personali per ulteriore finalità: sfide alla privacy*, in *Ciberspazio e diritto* 11.4, 2010; L. Piatti, *Blockchain, decentralizzazione e privacy: un nuovo approccio del diritto*, in *Ciberspazio e diritto: rivista internazionale di informatica giuridica*, 19, 1/2, 2018, 179-196.

DPCE online
ISSN: 2037-6677

avere accesso non autorizzato anche alla sua storia medica. L'anonimizzazione non è, poi, un processo permanente: esiste infatti il forte rischio che, in virtù dell'evoluzione tecnologica o di eventuali data breach che facciano trapelare informazioni rilevanti, i dati contenuti in un dataset possano essere re-identificati. Inoltre, essa non rende i dati inutilizzabili: questi potrebbero continuare a essere usati per specifiche finalità. In certi casi, alcuni dati personali potrebbero essere separati da un dataset, ad esempio si potrebbero anonimizzare i log di accesso a un sito mantenendo solo la data di accesso e la pagina, ma non le informazioni relative a chi ha eseguito tale accesso[9].

Alla luce delle tecniche sino ad ora brevemente illustrate, la Differential Privacy presenta chiari vantaggi, in particolare la sua resistenza ad un'ampia gamma di attacchi alla privacy, consentendo la raccolta e la pubblicazione di data patterns, proteggendo, al contempo, la privacy delle persone catturate in un set di dati. Altro punto di forza di tale tecnica è racchiuso nella trasparenza, avendo, in particolare, la capacità di condividere informazioni sulle analisi dei dati senza influire negativamente sulla privacy individuale. Di conseguenza, conserva un profilo di chiarezza, riguardo il margine di errore, di incertezza e di altre ulteriori variabili, che la maggior parte dei limiti di divulgazione statistica non consentono. A tale scopo, la privacy differenziale presenta un potenziale significativo per consentire un ampio accesso a dati che in precedenza non potevano essere condivisi, compresi dati sensibili come dati medici o finanziari. Può, altresì, aiutare a proteggere i soggetti dall'ampia gamma di potenziali pericoli relativi alla privacy dei dati, dagli assistenti vocali domestici ai dati delle transazioni fino ai dati del browser. Le violazioni della privacy sono diventate sempre più comuni e funzionalità avanzate come il riconoscimento facciale, i deepfake, la sorveglianza silente non hanno fatto altro che aumentare tali rischi.

Applicare la tecnica in esame ai dati prima di rilasciarli o condividerli può ridurre, inoltre, notevolmente la possibilità che eventuali dati identificabili possano essere raccolti da persone che leggono i risultati di un calcolo. Ciò contribuirebbe a rendere più semplice la collaborazione o il rilascio pubblico di informazioni, poiché le organizzazioni partecipanti possono essere certe di non aver condiviso informazioni sensibili.

La Differential Privacy potrebbe, inoltre, costituire un valido strumento contro il fenomeno dell'overexposure and extracting of consent, che determina la disperazione del consenso digitale[10], in una forma di volontariato obbligatorio. In altre parole, tali forme di patologie del consenso nascondono un sentimento paradossale di divario tra la preoccupazione degli utenti riguardo la tutela della loro privacy e il loro compulsivo e cieco atteggiamento che li induce a cliccare alle innumerevoli richieste sulla voce "Accetto", all'interno di menù a tendina sempre più insistenti e oscuri, in cui i singoli soggetti perdono inconsciamente, molto spesso, la loro privacy. Possiamo sentirci, in tal modo, sopraffatti dalle

[9]Cfr. *AEPD-EDPS joint paper on 10 misunderstandings related to anonymisation*, 2021, disponibile in https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings related_en.

[10] W. Hartzog, *Privacy's Blueprint*, Harvard University Press, 2018, 207 ff; A.E. Waldam, *Industry Unbound*, Cambridge University Press, 2021, 169 ff.

migliaia di richieste di accesso, autorizzazione e consenso ai nostri dati che acconsentiamo in maniera esausta e disperata. Altre volte, è invece, il design stesso che induce gli utenti digitali a cliccare su "Accetto", poiché pulsanti, segnali, disegni possono essere manipolati per poterli cliccare accidentalmente o annullarli, sottovalutandone l'importanza.

La Differential Privacy, con l'iniezione di rumore e l'anonimizzazione ab origine, consentirebbe di tutelare i dati personali da forme di progettazione di design in grado di estorcere il consenso digitale all'interno della nebbia di formulazioni confuse, hidden links, e menù annidati, e dalla relativa manipolazione indiretta delle grandi aziende, la cui sovraesposizione alle richieste di consenso indirizzate agli utenti determinano una progressiva erosione psicologica della percezione dell'importanza del valore della protezione dei dati personali e della relativa valutazione del rischio.

Nonostante le prospettive di miglioramento e di rafforzamento della privacy presentate dalla tecnica in esame, sono diversi gli interrogativi a cui rispondere, le sfide da affrontare nei prossimi anni per giungere ad un accurato perfezionamento e ad ulteriori garanzie di tale tecnica.

Ad esempio, molti approcci differenzialmente privati si applicano solo a dati o tipi di dati particolari, come dati univariati[11] o dati categorici, ma i dati del mondo reale spesso si presentano in altre forme.

Di conseguenza, alcuni ritengono opportuno ricorrere a "relaxed differentially private approaches"[12], sebbene questo si traduca in ulteriori rischi per la privacy[13].

Altro elemento di debolezza è rappresentato dalla mancanza di linee guida o strutture accertate su come impostare il c.d. parametro di perdita della privacy: senza la delineazione di standard ben precisi e la predisposizione di linee guida pratiche, è probabile che i dati vengano condivisi senza un'adeguata protezione della privacy o senza una quantità adeguata di utilità dei dati stessi.

Allo stesso modo, diversi approcci teorici richiedono notevoli risorse computazionali: ciò rende la Differential Privacy impegnativa e persino irrealizzabile per il curatore di dati medio con risorse computazionali limitate. Inoltre, tale tecnica si rivela essere particolarmente adatta per set di dati molto grandi, dove il rumore aggiuntivo non influisce in modo significativo sull'accuratezza o sull'utilità dei dati. Per set di dati più piccoli, invece, il compromesso tra privacy e utilità diventa più difficile e complesso da eseguire. I record-level data, in particolare, sono difficili da proteggere in modo significativo pur continuando a "lasciare i dati utili per scopi analitici non specificati"[14].

[11] I dati univariati sono dati che considerano una sola variabile nella sua unità e attraverso i c.d. indici univariati vengono utilizzati, pertanto, per descrivere singolarmente una variabile del dataset.

[12] Si tratta di metodologie di applicazione della Differential Privacy meno rigide e "rilassate", caratterizzate da una minore iniezione di rumore, inserito solamente nell'ultimo step della fase di training, ovvero la predizione del risultato. Cfr. L. Ninghui, *op.cit.*, pag. 30

[13] C. Dwork, N. Kohli, D. Mulligan, *Differential Privacy in Practice: Expose your Epsilons!* in *Journal of Privacy and Confidentiality*, 9(2), 2019.

[14] F.T. Wu, *Defining privacy and utility in data sets*, in *U. Colo. L. Rev.*, *84*, 1117, 2013.

DPCE online
ISSN: 2037-6677

La preoccupazione principale riguardo la privacy differenziale è il compromesso tra utilità dei dati e privacy individuale. Se il parametro di perdita di privacy è impostato in modo da favorire l'utilità, i vantaggi in termini di privacy vengono ridotti (nel sistema viene immesso meno "rumore"); se il parametro di perdita della privacy è, invece, impostato per favorire una privacy elevata, l'accuratezza e l'utilità del set di dati vengono diminuite (viene iniettato più "rumore" nel sistema). È importante che i decisori politici considerino i compromessi posti dalla DP al fine di contribuire a stabilire migliori pratiche e standard adeguati sull'utilizzo di tale pratica di tutela della privacy, soprattutto considerando la diversità dei casi d'uso organizzativi.

L'inserimento della Differential Privacy nel design permette, infatti, di proteggere da eventuali danni alla privacy prima ancora che si verificano, attraverso la sua singolare metodologia, sviluppando un approccio proattivo, che rende la DP uno strumento di valenza trasversale nei vari ordinamenti confinati ad un approccio reattivo, in cui le leggi di "reazione", incentrate causalmente sulla risposta ad una determinata condotta e ai relativi danni, non fungono, molto spesso, da deterrente. Infatti, alle volte, i danni alla privacy non possono essere rilevati nemmeno dalle vittime, ma ciò non significa che le stesse non siano state danneggiate; altre volte ancora, danni come il furto di identità possono essere così remoti da non poter essere collegati alla violazione. Inoltre, le aziende sono spesso disposte ad accettare il rischio di oltrepassare i limiti di ciò che sono autorizzati a raccogliere e condividere, poiché è probabile che una parte enorme di attività dubbie rimanga legalmente incontrastata. Alla luce di ciò, risulta fondamentale la ricerca di un binomio design-privacy, in cui la Differential Privacy possa costituire uno strumento in grado di fornire alle varie legislazioni una garanzia di protezione contro le innumerevoli forme di vulnerabilità e dipendenza tecnologica, una tecnica in grado di coniugare fiducia, oscurità "rumorosa" di protezione e autonomia riuscendo, in tal modo, a colmare al meglio il divario progettuale delle varie leggi sulla privacy e a costruire una privacy sostenibile in un mondo digitale.

3. Esempi applicativi e modelli di progettazione

Grandi aziende tecnologiche come Microsoft, Apple, Uber e Google hanno investito in modo significativo sulle applicazioni della Differential Privacy nei loro prodotti, aprendo a nuove prospettive futuristiche di trattamento della privacy.

Ad esempio, Google ha creato uno strumento per la privacy differenziale chiamato RAPPOR che tenta di applicare tale tecnica ai dati di utilizzo che raccoglie dal suo browser Chrome. Google ha reso open source una libreria per fornire una protezione differenziale della privacy a set di dati in più lingue e per una serie di statistiche riepilogative e descrittive. A tal proposito, di recente, Big G ha annunciato il progressivo spegnimento dei cookies, una graduale rivoluzione che per fine anno dovrebbe riguardare tutti gli utenti che utilizzano il browser Chrome, per quello che sarà uno dei più grandi cambiamenti nella storia di Internet. Su tale orizzonte di mutamenti futuristici, dal 2025 Google introdurrà un nuovo sistema di

tracciamento chiamato Privacy Sandbox, una sabbiera tecnologica volta a trasferire l'analisi delle attività degli utenti dallo strumento dei cookie di terze parti al Browser, riducendo il tracciamento cross-site e cross-app, contribuendo, in tal modo, a mantenere i contenuti e i servizi online liberi per tutti. Si tratta di un mutamento necessario in ottica di un miglioramento della tutela del consumatore e in particolare in tema privacy. Per questo motivo, all'interno dell'industry si stanno offrendo delle soluzioni di tracciamento alternative che permetteranno di garantire l'addressability pubblicitaria tutelando allo stesso tempo la privacy dell'utente.

Tuttavia, l'implementazione su larga scala della privacy differenziale fino ad oggi è stata effettuata dall'US Census Bureau nel suo censimento decennale del 2020[15]. Occorre previamente ricordare che l'Ufficio censimento degli Stati Uniti è tenuto ad effettuare, ogni dieci anni, una enumerazione effettiva di tutte le persone che vivono negli Stati Uniti, con l'obbligo di mantenere riservate le informazioni di identificazione personale per 72 anni. Il Title 13, U.S. Code, Section 9 , prevede, infatti, il compito per l'Ufficio di Presidenza di non «utilizzare le informazioni fornite ai sensi delle disposizioni di questo titolo per scopi diversi da quelli statistici per i quali sono fornite o effettuare qualsiasi pubblicazione in cui i dati forniti da un determinato istituto o individuo sotto questo titolo possano essere identificati o consentire a chiunque altro che non siano i funzionari giurati e i dipendenti del Dipartimento o ufficio o agenzia dello stesso di esaminare i singoli rapporti». Per garantire la severa applicazione di tale mandato, è stabilito che gli agenti che commettono violazioni, siano puniti con multa fino a 250.000 dollari e cinque anni di reclusione. Al di là di ogni obbligo legale, tuttavia, l'Ufficio è profondamente consapevole che la qualità e l'accuratezza dei suoi censimenti e sondaggi dipendono dalla sua capacità di mantenere la fiducia del pubblico; a tale scopo, i ricercatori all'interno dell'Ufficio hanno sottolineato come un impegno alla privacy costituisca una componente fondamentale della cultura istituzionale dell'Ufficio. La duplice esigenza di un conteggio accurato e della relativa protezione dei rispondenti e dei loro dati crea una tensione naturale: quanto più accurati (e quindi utilizzabili) sono i dati riportati, tanto più facile può essere l'identificazione delle singole risposte. Tuttavia, poiché i dati grezzi vengono alterati prima di essere riportati (per proteggere la riservatezza), meno utilizzabili risulteranno i dati resi pubblici. Dalla pubblicazione del censimento del 2010, il personale dell'Ufficio si è reso conto che i data analysts, grazie alla maggiore potenza di calcolo e alla crescita di altri database, come quelli utilizzati dai fornitori di dati commerciali, potevano prendere i numerosi dati prodotti dall'Ufficio e incrociarli tra loro o con fonti di dati esterne al punto che la privacy individuale e la relativa riservatezza, potrebbero essere compromesse. Sulla base, quindi, di tale pericolosa possibilità, l'Ufficio di Presidenza ha adottato dal 2020 la tecnica della Differential Privacy come

[15] M.B. Hawes, *Implementing differential privacy: Seven lessons from the 2020 United States Census*, in *Harvard Data Science Review* 2.2, 2020; C.T. Kenny, et al., *The use of differential privacy for census data and its impact on redistricting: The case of the 2020 US Census*, in *Science advances* 7.41, 2021; V.J. Hotz, J. Salvo, *A chronicle of the application of differential privacy to the 2020 Census*, in *Harvard Data Science Review*, Special Issue 2/2022; W.P. O'Hare, *Differential undercounts in the US census: who is missed?* London, 2019.

DPCE online
ISSN: 2037-6677

prospettiva di miglioramento per la garanzia della privacy. Attraverso tale tecnica, l'Ufficio ha dichiarato che la popolazione totale di ogni Stato sarà "as enumerated", ma che tutti gli altri livelli geografici - compresi i distretti congressuali fino alle township e ai blocchi di censimento - potrebbero avere qualche variazione rispetto ai dati grezzi. L'Ufficio del censimento definisce questa operazione come "iniezione di rumore" nei dati, indicando, altresì, che non verrà inserito alcun "white noise" nella popolazione totale dello Stato, ma è probabile che venga iniettato tale rumore distinguendo i diversi livelli geografici. Tale rivoluzione è stata sicuramente rassicurante per tutti coloro che erano preoccupati per la riservatezza dei dati del censimento, ma è stata accolta, allo stesso tempo, con scalpore da parte dei ricercatori di scienze sociali e dei politici che ripongono molto affidamento sui dati del censimento[16]. Tali dati, infatti, vengono utilizzati per la ripartizione della Camera dei rappresentanti, l'allocazione dei dollari dei contribuenti federali e la riorganizzazione dei distretti all'interno degli Stati, nonché per la ricerca, l'elaborazione di politiche basate sull'evidenza e il processo decisionale in materia di affari e investimenti.

La Nuova Zelanda, invece, ha sperimentato l'impiego della Differential Privacy come meccanismo innovativo per tutelare la privacy dei pazienti[17], garantendo che la condivisione dei dati sia bilanciata con la protezione della riservatezza delle informazioni sanitarie di ogni singolo individuo.

Le preoccupazioni relative alla salvaguardia dei dati sono, infatti, giustificate da numerose precedenti violazioni avvenute in Nuova Zelanda in relazione ai dati sanitari. In uno di questi esempi, un grafico del New Zealand Health, che riportava svariati dati riguardanti bambini trattati in ospedale per infezioni respiratorie, gastroenterite e pertosse, era stato visualizzato online dal 2005 al 2018. Il grafico era collegato ai dati di origine, inclusi nomi e data di nascita e i relativi risultati dei test, e aveva registrato 300 visualizzazioni prima che la violazione fosse identificata e i dati fossero, di conseguenza, rimossi. Un altro esempio di grave violazione riguardava i dati forniti da individui al Servizio trasfusionale della Croce Rossa tra il 2010 e il 2016. Le informazioni personali e sanitarie, compresi i dettagli sul c.d. "comportamento a rischio", erano state inserite in un modulo di richiesta online, il file contenente tutti questi dati era stato spostato su un dispositivo non sicuro e, pertanto, vi avevano avuto accesso diversi soggetti non autorizzati. Entrambi questi esempi riguardavano set di dati che contenevano identificatori personali. Vi sono stati anche casi in cui i set di dati sono stati condivisi con misure messe in atto per proteggere la riservatezza, come la rimozione di identificatori primari, ma si erano comunque verificate diverse violazioni. Ad esempio, il 1° agosto 2016, il

[16] F. Fioretto, P. Van Hentenryck, K. Zhu, *Differential privacy of hierarchical census data: an optimization approach*, in *Artificial Intelligence* 296, 2021; R. Gong, E.L. Groshen, S. Vadhan, *Harnessing the known unknowns: Differential privacy and the 2020 census*, in *Harvard Data Science Review* 6/2022.

[17] K. Weerasinghe, D. Pauleen, N. Taskin, S. Scahill, *Alignment of Big Data Perceptions Across Levels in Healthcare: The case of New Zealand*, in *Australasian Journal of Information Systems*, 27, 2023; A. Dyda, et al., *Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality*, in *Patterns* 2.12, 2021; C. Lin, et al., *Differential privacy preserving in big data analytics for connected health*, in *Journal of medical systems* 40, 2016.

Dipartimento della Salute neozelandese ha deciso di rendere pubblicamente i dati di fatturazione medica di 2.985.511 individui sul sito web data.gov.nz. I registri includevano dati del Medicare Benefits Scheme dal 1984 al 2014 e dati del Pharmaceutical Benefits Scheme dal 2003 al 2014, contenenti dati sanitari storici di circa il 10% della popolazione, inclusi dettagli sui servizi forniti da medici, patologi, diagnostici., imaging e servizi sanitari affini. Il set di dati è stato rilasciato per la ricerca nell'interesse pubblico e utilizzava identificatori univoci anonimizzati applicando al contempo numerose misure di riservatezza per impedire che le informazioni fossero identificabili, tra cui crittografia, perturbazione e aggregazione. Il set di dati è stato scaricato circa 1.500 volte mentre era disponibile al pubblico, eludendo ogni prospettiva di garanzia preposta.

Con la protezione aggiuntiva offerta dalla privacy differenziale, potrebbe essere possibile condividere informazioni più dettagliate sui modelli di salute e di malattia in un forum pubblico. Questa nuova prospettiva consentirebbe ai medici, agli operatori sanitari pubblici, ai ricercatori e al pubblico in generale di accedere a informazioni più tempestive e maggiormente complete sul quadro patologico, costituendo, pertanto, una soluzione di particolare importanza durante situazioni in rapida evoluzione come le emergenze sanitarie pubbliche, quali epidemie e pandemie. Ad esempio, durante la pandemia di COVID-19, una serie di indagini ha individuato la necessità di coinvolgere maggiormente i medici di base e gli altri operatori sanitari primari nella risposta all'emergenza, migliorando anche la loro capacità di accedere a dati dettagliati sull'epidemiologia e la diffusione della malattia. Tuttavia, la divulgazione di dettagli sui casi di COVID-19, come il luogo di esposizione e di contagio, i dati demografici e territoriali potrebbe potenzialmente consentire l'identificazione di individui che sono risultati positivi al tampone. La Differential Privacy potrebbe, invece, migliorare significativamente la protezione di tale tipologia di divulgazione dei dati. A questo proposito, la Nuova Zelanda ha sperimentato l'utilizzo di CRISPER[18], un sistema informativo in tempo reale per la preparazione e la risposta alle epidemie, sul solco dell'esempio dell'Australia, dove tale particolare applicazione di Differential Privacy è sorta. Questo sistema comprende diversi dashboards per visualizzare e interagire con i dati, compresi quelli relativi ai casi confermati e ai decessi, alla fonte di infezione, alla localizzazione dei contatti e ai test di laboratorio. CRISPER è progettato per utilizzare un algoritmo di privacy differenziale attraverso un motore di dati per proteggere i dati non disponibili pubblicamente (ad esempio, dati stratificati per età, sesso e comorbidità). L'accesso ai dati elencati è stata la sfida principale per il progetto CRISPER e il sistema attualmente utilizza raschiatori di dati e interfacce di programmazione delle applicazioni per analizzare tali dati da una serie di fonti pubbliche, come i siti web dei dipartimenti sanitari. L'uso della DP nel progetto CRISPER fornisce una prova di attuazione dell'applicazione di questo meccanismo innovativo, che potrebbe incoraggiare la condivisione dei dati sulla sanità

[18] E. Field, A. Dyda, et al., *Development of the COVID-19 Real-Time Information System for Preparedness and Epidemic Response (CRISPER), Australia. Frontiers* in *Public Health*, 9:753493, 2021; F. Dankar, K. Emam, *The application of differential privacy to health data*, in *ACM International Conference Proceeding Series*, 2012, 158–166.

DPCE online
ISSN: 2037-6677

pubblica in futuro, fornendo, di conseguenza, maggiore privacy e riservatezza. È bene sottolineare che, nei casi descritti, l'impiego della Differential Privacy nella tutela dei dati sanitari non deriva da un intervento normativo o regolamentare ma dall'azione di privati, tesi ad offrire soluzioni tempestive a problematiche di cogente risoluzione, mettendo a frutto le innovazioni tecnologiche sperimentate a tutela della privacy, in modo che essa non venga eclissata in nome di emergenze su larga scala.

4. Il Privacy Act 2020 della Nuova Zelanda in prospettiva di equilibrio

L'approccio singolare adottato dalla Nuova Zelanda nei confronti della protezione dei dati sanitari personali, rende tale Paese in grado di mantenere un modus operandi equilibrato tra il rispetto della rigidità normativa, rappresentata dal Privacy Act 2020, e l'apertura verso nuove prospettive futuristiche, quali la Differential Privacy. Questo atteggiamento rende la Nuova Zelanda in grado di porsi quale punto di equilibrio tra il modello europeo, confinato ai problemi legati all'anonimizzazione e alla pseudonimizzazione, e tra l'approccio statunitense, sicuramente propenso ad accogliere nuove prospettive avanguardiste ma carente di una solida e omogenea legislazione alla base. La Nuova Zelanda ha aggiornato la sua legge sulla protezione dei dati con il Privacy Act 2020, sostituendo, così, il precedente Privacy Act del 1993 e introducendo alcune modifiche significative per allinearsi agli standard e alle migliori pratiche internazionali. Si tratta di una legge che disciplina il modo in cui le organizzazioni e le aziende possono raccogliere, archiviare, utilizzare e condividere le informazioni personali degli individui in Nuova Zelanda, mirando, altresì, a proteggere i diritti alla privacy dei singoli soggetti e a garantire che le loro informazioni personali siano gestite in modo equo, trasparente e sicuro[19]. Con l'introduzione del nuovo Privacy Act, la Nuova Zelanda presenta una maggiore propensione per adottare e ricorrere alla Differential Privacy. Infatti, seppur non vi sia un preciso richiamo, all'interno della nuova legislazione, che disciplini il ricorso a tale tecnica, è possibile ricavare dall'art. 22 del Privacy Act l'importanza di adottare misure per evitare rischi per la privacy prima che si verifichino, promuovendo, pertanto, un approccio più proattivo alla sicurezza dei dati, tipico della metodologia dell'iniezione di rumore prevista dalla DP.

A differenza dei sette principi su cui si basa il GDPR, il Privacy Act neozelandese presenta tredici punti cardine: finalità e liceità della raccolta dei dati personali; raccolta di informazioni personali preferibilmente presso l'interessato; obblighi di informativa e consenso per la raccolta di informazioni personali; misure di sicurezza contro la perdita, l'uso improprio e la divulgazione; diritti di accesso e rettifica; garanzia dell'accuratezza e della qualità delle informazioni prima di utilizzarle e divulgarle; limiti all'utilizzo e alla conservazione di tali informazioni personali in base alla necessità; condizioni per la divulgazione e il trasferimento di dati personali,

[19] S. Penk, N. Chamberlain, *Privacy law in New Zealand*, Wellington, 2023; Y. Dong, *Privacy Act 2020*, in *Auckland UL Rev.* 26, 2020; M. Bartlett, *Beyond privacy: Protecting data interests in the age of artificial intelligence*, in *Law, Technology and Humans* 3.1, 2021.

in particolare all'estero; limitazione dell'uso di identificatori univoci. Tuttavia il GDPR e il Privacy Act 2020 condividono alcuni obiettivi e caratteristiche comuni, quali il rafforzamento della protezione dei diritti alla privacy dei singoli individui, garantendo loro un maggiore ed efficace controllo sulle informazioni personali, la semplificazione del contesto normativo per le imprese e per le organizzazioni che gestiscono informazioni personali a livello transfrontaliero, la promozione della trasparenza, della responsabilità e della sicurezza nel trattamento delle informazioni personali, la predisposizione di meccanismi per la supervisione, l'applicazione e il ricorso.

Vi sono, però, tra le due regolamentazioni delle differenze sostanziali notevoli. Ad esempio, la legge neozelandese sulla privacy non definisce i concetti di dati anonimizzati o de-identificati. Tuttavia, consente l'uso di informazioni personali per scopi secondari e la relativa divulgazione ad un soggetto terzo o organizzazione se sono «utilizzate in una forma in cui l'individuo interessato non possa essere identificato[20]». Non è chiaro se ciò richieda semplicemente di privare i dati di identificatori diretti o se si tratti di qualcosa di più rigoroso. Poiché le informazioni personali sono definite come "informazioni su un individuo identificabile", è logico che "non identificato" e "non identificabile" abbiano un significato diverso, perché se l'individuo non fosse identificabile, la legge sulla privacy non si applicherebbe e non sarebbe necessaria alcuna eccezione specifica che consenta usi secondari o divulgazione. Il Privacy Act 2020 non richiede, inoltre, il consenso (definito dalla legge "authorization") come base per la raccolta o l'utilizzo di informazioni personali, nemmeno per le informazioni sensibili, anche se il Commissario per la privacy afferma che alle informazioni personali sensibili si applica uno standard di cura più elevato[21]: è anche una delle eccezioni alla regola secondo cui le informazioni devono essere raccolte dall'individuo stesso, piuttosto che da terzi, alla regola secondo cui le informazioni personali non possono essere utilizzate per scopi diversi da quelli per cui sono state raccolte e alla regola secondo cui le informazioni personali non possono essere divulgate a terzi e a una giurisdizione straniera. Il GDPR richiede, invece, il consenso - libero, informato, specifico, inequivocabile ed esplicito per i dati sensibili[22] - come una delle sei basi giuridiche per il trattamento dei dati personali. Altra importante differenza tra le due regolamentazioni risiede nel fatto che mentre il Privacy Act 2020 conferisce limitatamente alle persone il diritto di accedere e correggere le proprie informazioni personali, il GDPR fornisce, invece, ai singoli individui diritti più ampi, come il diritto alla cancellazione (noto anche come diritto all'oblio), il diritto alla limitazione del trattamento, il diritto alla portabilità dei dati e il diritto di opporsi al trattamento. Altro elemento di differenza di fondamentale importanza risiede nel fatto che mentre il GDPR, all'art.22 affronta la delicata questione del trattamento automatizzato, sostenendo che l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo

[20] Cfr. art. 22 comma 3 del Privacy Act 2020.
[21] Cfr. art. 22 del Privacy Act 2020.
[22] Cfr. art. 7 del GDPR.

DPCE online
ISSN: 2037-6677

significativamente sulla sua persona, il Privacy Act non effettua alcuna menzione in merito.

Alla luce di tali considerazioni sembra emergere da parte della Nuova Zelanda un approccio alla disciplina della privacy meno rigido rispetto a quello presentato dalla General Data Protection Regulation, particolarmente singolare nella sua specialità, infatti pur ispirandosi marginalmente alle disposizioni del GDPR, e potendo quindi contribuire all'interoperabilità dei regimi di privacy in tutto il mondo, rimane fedele alla storia decennale della propria Nazione di regolamentazione della privacy e della protezione dei dati, con uno sguardo attento alle nuove sfide presentate delle forze rivoluzionarie digitali. Totalmente diverso risulta essere l'approccio statunitense sul profilo di regolamentazione della protezione dei dati personali. Negli Usa, infatti, al centro del sistema vi è l'autonomia dei privati e la libertà individuale: l'approccio è, quindi, di tipo autoregolamentante, utilitaristico, oltre che settoriale, laddove la privacy è tutelata solo nell'ambito delle pratiche commerciali, tramite l'equilibrio del mercato[23]. Pertanto, negli Usa la tutela della privacy è attribuita principalmente alla Commissione per il Commercio Federale (FTC, Federal Trade Commission), in una sostanziale estensione della tutela del consumatore e della legittimità delle pratiche commerciali. L'approccio americano è sicuramente più efficace ed adattabile alle mutazioni tecnologiche, ma in compenso finisce per far diventare la privacy un bene economico da poter scambiare all'interno di un ampio mercato dei dati personali, così svalorizzandone inesorabilmente l'aspetto individuale. La tutela è, quindi, indiretta, in quanto basata prevalentemente sulle norme a tutela dei consumatori, per cui il regime dominante di regolamentazione della privacy è noto come notice and choice. Tuttavia l'informativa e il consenso si rivelano inadeguati sia nell'informare le persone sia nell'offrire loro la tutela della privacy. Pochi, infatti, leggono le politiche lunghe e legalitarie, che sembrano estendersi per pagine contorte offuscate e dai margini labili e opachi, allegate ai prodotti, alle app e ai servizi che utilizzano ogni giorno. Anche se le informative fossero complete e di facile comprensione, è praticamente impossibile per gli individui fornire un consenso significativo. L'Europa e altri Paesi hanno incoraggiato gli Stati Uniti ad adottare una legge simile al Regolamento generale sulla protezione dei dati dell'UE. Tuttavia, un "GDPR statunitense" sembra essere destinato a soffrire delle stesse patologie del consenso sopra descritte, mentre un "Privacy Act statunitense" dal profilo neozelandese sembrerebbe mitigare l'abuso eccessivo del consenso, ricorrendo ad una forma di "autorizzazione", che si ponga al riparo dalle derive del consenso fittizio, fabbricato, disperato o forzato[24], riducendo, di conseguenza, al minimo i rischi e i danni della economia dell'informazione, in una prospettiva di perseguimento di vantaggi

[23] S.S. Bakare, et al., *Data privacy laws and compliance: a comparative review of the EU GDPR and USA regulations*, in *Computer Science & IT Research Journal* 5.3, 2024; A. Somma, *Gli Stati Uniti e il loro diritto*, Torino, 2023; D.J. Solove, P.M. Schwartz, *Information privacy law*, Boston, 2020; S.G. Jamison, *Creating a National Data Privacy Law for the United States*, in *Cybaris Intell. Prop. L. Rev.* 10, 2019; J.A. Rothchild, *Against notice and choice: The manifest failure of the proceduralist paradigm to protect privacy online (or anywhere else)*, in *Clev. St. L. Rev.* 66, 2017.
[24] W. Hartzog, *op.cit.*, 207 ff.

ottenibili in modo sostenibile, etico e progressivo. Inoltre, l'esempio della legislazione della Nuova Zelanda consentirebbe agli Stati Uniti di porre soluzione al problema della "theory of privacy nicks"[25], secondo cui i legislatori hanno assunto sistematicamente un atteggiamento di normalizzazione della sorveglianza ignorando le diminuzioni della privacy più piccole, frequenti e marginali, perseguendo giuridicamente solo le violazioni della privacy più urgenti e gravi, le c.d. privacy chops, alludendo al colpo di lama rapido e affilato. L'adozione della DP consentirebbe, altresì, di porre un freno ad una sorveglianza silente e assidua, proteggendo la sicurezza dei dati personali, indipendentemente, dalle lame più o meno affilate dei nicks o dei chops.

5. Conclusioni

Sulla base delle considerazioni svolte, la Differential Privacy potrebbe in futuro aprire una nuova frontiera nella tutela dei dati personali, costituendo una metodologia particolarmente rilevante in contesti in cui la raccolta e l'analisi dei dati sono sensibili dal punto di vista della privacy, come nel campo della sanità, delle politiche pubbliche o delle ricerche di mercato. Sono numerose, infatti, le violazioni dei dati che minacciano governi, organizzazioni e aziende. Allo stesso tempo, le odierne applicazioni di machine learning si basano su tecniche di apprendimento che richiedono grandi quantità di dati di addestramento, spesso provenienti da individui. La divulgazione impropria di tali dati, in qualsiasi forma, può causare molti problemi sia all'individuo che all'organizzazione e, nei casi più gravi, può portare a forme di responsabilità civile o penale.

In secondo luogo, la Differential Privacy potrebbe superare le criticità tipiche delle tecniche di anonimizzazione e pseudonimizzazione, le quali spesso si caratterizzano per essere un processo non permanente, che non assicura l'inutilizzabilità del dato, mantenendo un rischio elevato di identificabilità. Inoltre, l'avvento della "privacy diversa" rispecchia, a mio avviso, l'evoluzione del concetto stesso di privacy nell'era dell'innovazione tecnologica: dalla concezione di riservatezza intesa come diritto di essere lasciati soli, the right to be alone, di circa un secolo fa, si è arrivati alla codifica della protezione dei dati personali come un habeas data nella carta di Nizza. Ad assumere rilievo centrale non è più la sola tutela dell'individuo nella sua sfera privata, ma anche il diritto all'autodeterminazione informativa, cioè alla possibilità da parte dell'interessato di seguire il completo ciclo di vita del dato.

Tuttavia, è necessario procedere ad un maggiore perfezionamento della tecnica in questione, in modo da risolvere i punti di debolezza e le criticità che impediscono una applicazione più accurata e diffusa e che ne limitano il raggiungimento al di fuori delle comunità scientifiche. Similmente, è auspicabile un maggior intervento normativo, soprattutto a livello europeo, in grado di regolamentare l'applicazione della Differential Privacy, i relativi standard e le modalità di ricorso, in modo da realizzare una progressiva e controllata apertura verso tale nuova frontiera. Pertanto,

[25] W. Hartzog, E. Selinger, Evan, J. Gunawan, *Privacy Nicks: How the Law Normalizes Surveillance*, in *101 Washington University Law Review* 717, 2024.

DPCE online
ISSN: 2037-6677

l'Europa e gli USA dovrebbero guardare all'esempio neozelandese che getta le prime basi verso approcci innovativi per la tutela della privacy, al passo con le nuove realtà inaugurate dall'innovazione tecnologica e le sfide aperte dell'Intelligenza Artificiale. Il Privacy Act della Nuova Zelanda potrebbe essere, a sorpresa, il nuovo faro per il GDPR e la situazione statunitense.

Roberta Nobile
Dipartimento di Giurisprudenza Università di Catania
Università Campus Bio-Medico di Roma
robertanobile110@gmail.com