Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale
In Marketing e Comunicazione

Tesi di Laurea

La Percezione della Privacy su Twitter

Un'analisi del sentimento attraverso il confronto fra tweet

Relatore
Ch. Prof. Emanuele Aliverti

Laureanda
Claudia Borgogelli
Matricola 889592

Anno Accademico
2022/2023

Alla mia famiglia

INDICE

INTRODUZIONE

"La privacy è più di una questione tecnologica; è una questione di libertà e di dignità umana."
(Sheryl Sandberg)

Privacy e tutela dei dati personali. Due concetti ritenuti simili, ma diversi tra loro, divenuti ormai di scottante attualità e di comune interesse tra i cittadini. Nell'era digitale in cui viviamo, infatti, la tutela della *privacy*, unitamente alla protezione dei dati personali, ha assunto sempre più rilevanza, soprattutto alla luce dell'enorme quantità di dati generati, condivisi e scambiati su base quotidiana attraverso la rete internet, le piattaforme online ed i *social media*. Tali dati, sono diventati uno strumento fondamentale per le imprese, in special modo nel campo del marketing, ormai legato indissolubilmente al mondo del cliente. In particolare, in virtù di una conoscenza più approfondita dei clienti, l'analisi dei *social media data* consente la realizzazione di strategie di maggior successo ed efficacia. Tuttavia, il loro utilizzo può sollevare nei consumatori diversi timori e preoccupazioni legati, appunto, alla tutela della *privacy* e dei propri dati personali. Un'integrazione responsabile di quest'ultimi nel marketing rappresenta, quindi, la chiave per mantenere il delicato equilibrio tra consumatori e aziende.

Il presente elaborato, prendendo le mosse da questa tematica, si propone di comprendere come una percezione positiva o negativa dei consumatori relativamente alla propria *privacy* ed alla tutela dei dati personali, possa influenzarli nelle relazioni con l'azienda e nel rilascio di dati ed informazioni. A tale scopo, è stata individuata la piattaforma Twitter quale fonte dei dati e la cosiddetta analisi del sentimento quale strumento operativo. In particolare, la scelta di utilizzare la *sentiment analysis* deriva dal desiderio di dimostrare l'importanza strategica per le aziende di conoscere, non solo i gusti e le preferenze dei clienti, ma anche le loro emozioni e sensazioni, i loro pensieri ed opinioni, in quanto i clienti stessi rappresentano la loro principale risorsa a disposizione.

Il lavoro si articola in tre capitoli. Il primo contiene un approfondimento in merito alla genesi ed allo sviluppo del concetto di *privacy*, alle sue differenze con la tutela dei dati personali, alla correlazione con la normativa europea (GDPR) e, infine, al rapporto con la società digitale attuale. Nel secondo capitolo viene introdotto il tema dei *social media*, con un focus specifico su come i dati provenienti da quest'ultimi, denominati *social media data*, possano rappresentare una fonte di dati per la realizzazione di studi ed analisi. Inoltre, considerando l'ampia diffusione

ed il ruolo pervasivo che i *social media* hanno assunto negli ultimi decenni, è stato doveroso inserire una riflessione sui problemi legati alla *privacy* e sull'aspetto etico connesso al loro utilizzo. Infine, l'ultimo capitolo è dedicato all'analisi effettuata sui dati di Twitter. La prima parte concerne l'analisi esplorativa delle parole più frequenti, degli *hashtag*, dei *like* e dei *retweet* al fine di capire come si articola su Twitter il dibattito sul tema della *privacy*. La seconda parte, invece, si concentra sull'analisi del sentimento, utilizzando il software 'R' per comprendere le emozioni e gli stati d'animo degli utenti sul tema della tutela della *privacy* e dei dati personali.

In sintesi, quanto è importante garantire la *privacy* e la sicurezza dei dati personali dei propri clienti? E poi, è veramente fondamentale l'equilibrio tra l'utilizzo dei *social media data* ed il rispetto del cliente/utente?

1. PRIVACY E PROTEZIONE DATI PERSONALI

"Affermare che non si è interessati al diritto alla privacy perché non si ha nulla da nascondere è come dire che non si è interessati alla libertà di parola perché non si ha nulla da dire." (Edward Snowden)

1.1- CONCETTO DI PRIVACY: ORIGINE E SVILUPPI

Al giorno d'oggi la tematica della *privacy* ricopre un ruolo di primaria importanza nelle nostre vite, soprattutto alla luce dell'ampia diffusione di internet, degli strumenti digitali e delle varie piattaforme tecnologiche. Dare una definizione univoca di *privacy* è complesso, in quanto come scrive Niger, *"la nozione di privacy non è una nozione unificante. Non è, cioè, un concetto che esprime esigenze uniformemente e coerentemente diffuse nella storia e nella collettività*" (Niger, 2006). È un concetto, quindi, che risulta in continuo mutamento e strettamente legato al contesto politico, sociale e culturale di ogni periodo storico (Solove, 2015). Una delle definizioni più chiare e attuali di tale concetto è stata formulata da Alan Westin, professore di diritto pubblico presso la Columbia University e uno dei maggiori esperti statunitensi del XX secolo in materia, il quale sosteneva che "*La privacy, dunque, riconosciuta pienamente come diritto, è anche potere, che scaturisce da un insindacabile atto di volontà. E' una pretesa, legittima, che ogni individuo ha di decidere in che misura e con che modalità, vuole condividere una parte di sé con gli altri…"* (Westin, 1967).

Ripercorrendo l'evoluzione storico-temporale del termine *privacy* è possibile risalire alle sue origini e comprendere pienamente la sua natura poliedrica.
Il concetto di *privacy* affonda le proprie radici indietro nel tempo; fin dall'antichità, infatti, vi era la necessità di proteggere e tutelare se stessi e i propri spazi da intrusioni esterne.
Già nell'Antica Grecia era possibile riscontrare, all'interno di trattati filosofici di rilievo, la tematica della riservatezza. Nell'opera 'La Politica', Aristotele teorizzò la separazione fra la sfera pubblica e privata degli individui, associate rispettivamente all'attività politica e alla vita familiare domestica (Aristotele, *La Politica*). All'epoca, infatti, era fondamentale che i cittadini maschi partecipassero alla vita pubblica, considerata quasi un vero e proprio dovere da rispettare (Arendt, 2001), così come era fondamentale che ogni cittadino avesse una sfera privata, limitata strettamente a soddisfare i propri bisogni e le proprie necessità (Niger, 2006).

Successivamente, con l'avvento della città-stato, la dimensione politica e quindi pubblica della vita di un individuo assunse un ruolo sempre più preponderante tanto che, tutti coloro che sceglievano di condurre una vita prevalentemente privata, non potevano essere considerati pienamente umani (Arendt, 2001). Tuttavia, nonostante l'aspetto sociale fosse estremamente importante, i confini della proprietà privata di un soggetto erano ritenuti sacri. La casa rappresentava l'origine delle attività economiche e, senza di essa, gli uomini non potevano partecipare agli affari cittadini.

In opposizione alla vivace vita pubblica della polis greca si trova la convivialità familiare tipica della società feudale. In epoca medievale, infatti, la connotazione del termine 'privato' divenne sinonimo di familiare. La vita privata era caratterizzata da quel senso di appartenenza, sicurezza e protezione propri di una famiglia. Non vi era spazio per l'individualità poiché gli individui erano organizzati in gruppi, uniti sulla base di una fiducia reciproca e chiunque provasse ad isolarsi veniva visto con sospetto e spesso considerato un mago o una strega. È solo con la disgregazione della società feudale che iniziò a diffondersi tra la borghesia un nuovo bisogno di intimità, funzionale al riconoscimento della propria identità all'interno della società (Rodotà, 1974).

Cominciò, così, ad affermarsi una connotazione del concetto di *privacy* più vicina a quella che vige ai giorni nostri. Infatti, con l'avvento dell'età moderna, si assiste alla nascita del concetto di *privacy* moderna, la quale "*fu scoperta come l'opposto non della sfera politica ma di quella sociale, alla quale è di conseguenza più strettamente e autenticamente connessa*" (Arendt, 2001). L'uomo valoroso non è più soltanto colui in grado di eccellere nel discorso e pronto all'azione, come avveniva nell'antica Grecia, ma soprattutto colui che lavora e si impegna a favore della comunità. Il lavoro e la produttività non sono più relegati alla dimensione domestica, ma diventano parte integrante della vita pubblica. Il fatto quindi, di essere molto attivo ed esposto a livello sociale, spinge l'uomo moderno a voler adottare una certa riservatezza per quanto riguarda i suoi aspetti privati.

La prima vera teorizzazione del concetto di *privacy* risale al 1890, anno in cui due giuristi statunitensi, Samuel Warren e Louis Brandeis, pubblicarono sulle pagine del "Harvard Law Review" un saggio intitolato '*The Right to Privacy'* (Warren, Brandeis, 1890). In quegli anni, la stampa andava progressivamente affermandosi come metodo di diffusione di notizie, attraverso articoli, e soprattutto fotografie, che potevano potenzialmente mettere in luce aspetti e dettagli della vita privata dell'alta società. Tale aspetto spinse Warren e Brandeis a teorizzare la necessità di un diritto che proteggesse la sfera personale dell'uomo, definito poi in chiave negativa come '*the right to be let alone*' ossia, non tanto come il diritto di proteggere i propri

dati ed informazioni personali da altri, ma come il diritto di privare gli altri delle proprie informazioni personali (Fioriglio, 2008) evitando, quindi, un'intrusione dell'intimità unitamente all'esigenza di vivere la propria vita privata con un senso di tranquillità.

'The Right to Privacy' rappresenta il primo scritto di carattere giuridico che riconosce e tutela un diritto alla *privacy*. Precedentemente, infatti, la giurisdizione tendeva a ricondurre le esigenze di tutela della vita privata a diritti legati alla reputazione e all'onore. Attraverso il loro operato, Warren e Brandeis hanno contribuito a portare all'attenzione del legislatore la necessità di protezione della dimensione privata, ponendo in primo piano l'uomo e il valore che l'intimità ha per lui (Miglietti, 2014).

1.1.1 - EVOLUZIONE NEGLI STATI UNITI

Dalla pubblicazione di *'The Right to Privacy'* fino agli anni '60 del novecento, l'evoluzione giuridica statunitense in materia di *privacy* fu pressoché inesistente. Nel 1960 la Corte Suprema emise una serie di sentenze che riconobbero la *privacy* e la tutelarono sia nella sfera pubblica, sia in quella privata. A seguito di tali sentenze e della pubblicazione del saggio '*Privacy'*, in cui William Prosser teorizzava una concezione pluralistica della *privacy*[1], si riaccese il dibattito dottrinale. Tale teoria, che si contrapponeva a quella unitaria ed individualistica di Warren e Brandeis, non fu esente da critiche. In particolare, Edward Blounstein si pose fortemente in opposizione a Prosser sostenendo la necessità di concepire la *privacy* da un punto di vista unitario, in quanto rappresentava un valore essenziale per l'uomo e un diritto meritevole di essere tutelato (Miglietti, 2014).

Parallelamente al dibattito giuridico, a livello sociale il continuo sviluppo delle tecnologie aveva favorito l'ingresso della società moderna nell'era dell'informazione, migliorando la vita dei cittadini, ma contribuendo, al contempo, a rendere la loro esistenza, le loro vicende e la loro identità sempre più pubblica. Il crescente utilizzo di nuove tecnologie informatiche da parte di istituzioni ed imprese fece emergere la necessità per i cittadini di una maggiore trasparenza da parte della pubblica amministrazione, che si concretizzò nel 1966 con il *Freedom of Information Act* (FOIA). Il FOIA consentiva ai cittadini di accedere sia a tutte le informazioni riguardanti gli enti pubblici, sia a tutte quelle detenute da quest'ultimi. In aggiunta ad esso, nel 1974 venne emanato il *Privacy Act*, una legge federale che rappresenta ancora oggi un testo di riferimento in materia di *privacy*. Tale disciplina era pensata per regolare i rapporti tra le istituzioni e gli

[1] La concezione della *privacy* elaborata da William Prosser era di tipo pluralistico poiché egli sosteneva che da una violazione della *privacy* non scaturisse l'imputazione di un solo illecito ma di ben quattro, differenti tra loro: *inclusion upon seclusion, public discolsure of private facts, false light in public eye e appropriation.*

individui, obbligando le istituzioni a mantenere confidenziali i dati dei privati (Lugaresi, 2000). FOIA e *Privacy Act* rappresentano, quindi, due forme di tutela della *privacy* equilibrate, in quanto bilanciano la necessità e il diritto dei cittadini alla trasparenza e il diritto a proteggere la propria intimità da altri.

In particolare, tali leggi tutelano i dati dei privati solo in ambito di attività economiche, laddove si configurano i rischi per il cittadino "consumatore": la *privacy* viene considerata come diritto del consumatore e non dell'uomo in quanto tale.

La natura settoriale del sistema di tutela della *privacy* è stata messa a dura prova dall'attacco terroristico alle torri gemelle dell'11 settembre 2001, che ha contribuito a rendere le libertà individuali di ognuno degli strumenti per combattere la lotta al terrorismo. Dopo l'attentato, infatti, il governo americano ha messo a punto tutta una serie di cambiamenti in ambito di sorveglianza e ha aumentato vertiginosamente i controlli su Internet e su tutti gli altri strumenti digitali. Sarà solo nel 2013, attraverso le rivelazioni fatte da Edward Snowden, che il mondo scoprirà l'entità dei programmi di sorveglianza istituiti dalle agenzie di intelligence statunitensi. Purtroppo, già allora era riscontrabile uno dei grandi paradossi che lo sviluppo tecnologico ha portato con sé, ovvero il contrasto tra la crescente trasparenza degli individui, ottenuta attraverso il potere della condivisione e le reti Internet e l'operato sempre più criptico di governi e istituzioni. Ne è un esempio, sempre a seguito dei fatti dell'11 Settembre, la cancellazione dal web di informazioni istituzionali relative a molti Stati americani, operata dagli stessi per timore che potessero essere utilizzate da organizzazioni terroristiche per colpire nuovamente (Iaselli, Gorla, 2015).

È chiaro, quindi, come l'attentato alle Torri gemelle abbia avuto degli effetti disastrosi per le libertà individuali, non solo negli Stati Uniti, ma a ruota anche in tutti gli altri paesi occidentali. Tali effetti hanno contribuito ad influenzare e plasmare lo sviluppo tecnologico dello specifico settore che, nato con l'obiettivo di facilitare e migliorare la vita di tutti, in realtà è rimasto vittima di un meccanismo non ottimizzato fin dall'inizio.

1.1.2 - EVOLUZIONE IN EUROPA

La tutela in materia di *privacy* e protezione dati personali è considerata, ad oggi, uno dei caratteri distintivi del sistema giuridico europeo, il quale identifica le due discipline sopracitate come diritti fondamentali dell'individuo. Tale approccio europeo di tutela, prende le distanze da quello statunitense per via della considerazione di cui godono *privacy* e protezione dati personali, considerati da quest'ultimo, invece, come diritti del consumatore e non fondamentali per l'uomo in sé. Il sistema giuridico europeo nasce in origine come Comunità Economica

Europea (CEE) con l'obiettivo di creare integrazione economica tra gli Stati membri, attraverso la libera circolazione di persone, merci, servizi e capitali. All'interno di tale contesto riservatezza e *privacy* erano due tematiche riconducibili alla sfera dei diritti umani, per cui non godevano di specifiche disposizioni normative; il vuoto normativo è stato colmato successivamente attraverso la giurisprudenza delle Corti[2]. Così come avvenuto nel continente americano, anche in Europa l'evoluzione del diritto alla *privacy* ha incontrato molteplici ostacoli e difficoltà. In particolare, un primo approccio al tema risale al 1950, anno in cui venne firmata a Roma la *"Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali"* la quale sancisce, all'articolo 8, il rispetto della sfera privata di ciascuno, attraverso il divieto di ingerenza delle autorità pubbliche all'esercizio delle libertà individuali[3]. La convenzione non presenta, tuttavia, una forma specifica di disciplina della *privacy* e di tutela dei dati personali, i quali saranno regolati più accuratamente dal Consiglio d'Europa, nel 1981, con la *"Convenzione n. 108 sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale"* e successivamente dalla Carta di Nizza nel 2000. Al Consiglio d'Europa è riconducibile, quindi, l'introduzione per la prima volta di maggiori garanzie a favore di quei dati personali catalogati come "speciali", i cosiddetti dati sensibili. Inoltre, si deve alla Convenzione n. 108 l'introduzione del principio della protezione "equivalente", che consente il trasferimento di dati personali tra due Stati aderenti alla convenzione, solo nel caso in cui lo Stato destinatario fornisca le stesse garanzie e tutele adottate dallo Stato mittente[4].

Il 1993 rappresenta un punto di svolta. Si assiste, infatti, alla nascita dell'Unione Europea per mezzo del trattato di Maastricht, che sancisce anche la consacrazione formale dei diritti umani all'interno dei testi normativi comunitari (Miglietti, 2014). A tale proposito, la disciplina più significativa si ritrova nella Carta dei Diritti Fondamentali dell'UE (Carta di Nizza) del 2000, la quale all'articolo 8, garantisce ad ogni individuo il diritto di protezione dei dati personali che lo riguardano, definendo inoltre modalità e limiti del trattamento e prevedendo l'istituzione di

[2] Si fa riferimento alla Corte Europea dei Diritti dell'Uomo e alla Corte di Giustizia dell'Unione Europea.

[3] La Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali all'art. 8 recita come segue: *"1. Ogni persona ha il diritto al rispetto della propria vita privata e familiare, del proprio domicilio e della propria corrispondenza.* 2. *Non può esservi ingerenza di una autorità pubblica nell'esercizio di tale diritto a meno che tale ingerenza sia prevista dalla legge e costituisca una misura che, in una società democratica, è necessaria alla sicurezza nazionale, alla pubblica sicurezza, al benessere economico del paese, alla difesa dell'ordine e alla prevenzione dei reati, alla protezione della salute o della morale, o alla protezione dei diritti e delle libertà altrui*".

[4] Principio che verrà riformulato successivamente dall'Unione Europea nella prima direttiva emanata in materia di trattamento dei dati personali.

un'autorità indipendente preposta al controllo[5]. Con l'assunzione di importanza a livello istituzionale e di validità giuridica da parte del concetto di *privacy*, si diffonde l'espressione *"data protection"*, ossia protezione dei dati, che comporta la mutazione dell'originario diritto alla riservatezza in un diritto al "controllo dei dati".

La direttiva 95/46/CE del Parlamento Europeo e del Consiglio dell'Unione Europea (detta *"Data Protection Directive"* o anche direttiva "madre") rappresenta il primo provvedimento per l'introduzione di una normativa specifica sul trattamento dei dati personali negli Stati membri dell'Unione. Essa consente al legislatore europeo di fornire un'accurata definizione del concetto di dati personali concretizzando, inoltre, il nuovo profilo assunto dalla *privacy* nella tutela delle persone fisiche relativamente al trattamento dei dati personali. Obiettivo della direttiva era favorire un equilibrio tra il diritto alla vita privata e la libera circolazione dei dati fra gli Stati membri, purché quest'ultima non fosse in contrasto con i diritti fondamentali dell'uomo. L'innovatività di tale provvedimento, rispetto ai precedenti interventi in materia, consiste nel porre al centro dell'attività di trattamento dei dati gli individui e la loro vita privata (Miglietti, 2014). L'individuo, infatti, gode del diritto di avere informazioni sul modo in cui i propri dati personali vengono trattati ed eventualmente opporsi, ma anche il diritto di accedere a tali dati, rettificarli o cancellarli. Al fine di garantire la corretta applicazione della normativa e tutelare così ciascun individuo, la direttiva 95/46/CE ha imposto agli Stati membri una serie di principi, regole e misure di sicurezza a cui adattarsi e, inoltre, ha istituito un'apposita autorità di controllo con potere investigativo, giudiziario e di intervento in casi di violazioni delle disposizioni[6]. Attraverso la citata direttiva, il legislatore intendeva creare un nucleo di regole, principi e criteri che garantissero sul territorio dell'Unione una produzione omogenea dei dati personali. Tuttavia, non è stato possibile realizzare a livello europeo l'armonizzazione normativa auspicata dal legislatore in materia di *privacy*, a causa delle differenze nell'interpretazione, trasposizione e, conseguentemente, nell'adattamento degli Stati membri alle norme comunitarie. Successivamente, la direttiva "madre" subì ulteriori modifiche finalizzate ad adattarla ai continui sviluppi in ambito tecnologico, che minacciavano di mettere a rischio la vita privata di ciascun individuo e la sicurezza dei suoi dati personali. Le norme

[5] La Carta dei diritti fondamentali dell'UE all'art. 8 recita come segue: *"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"*.

[6] La direttiva "madre" ha favorito, in Italia, l'adozione della I n. 675/96 che ha portato all'istituzione della figura del Garante della privacy, oltre che al trattamento dei dati personali così come previsto dall'Unione Europea.

comunitarie in materia di *privacy* rimasero piuttosto obsolete, dimostrandosi inadatte a garantire i necessari livelli di protezione. Sarà solo nel gennaio del 2012 che la Commissione Europea si pronuncerà sul tema, proponendo una riforma sulla protezione dei dati che favorisca lo sviluppo di un'economia digitale (Mensi, Falletta, 2015) tenendo conto della *privacy online*, necessità sviluppatasi a seguito della diffusione di Internet e dei *social network*. Dopo un lungo periodo di negoziati, durato quasi quattro anni, è stato raggiunto nel 2015 un accordo tra Parlamento, Commissione e Consiglio dell'Unione Europea relativamente ai provvedimenti da attuare. Nello specifico, le proposte legislative presentate, adottate dal Parlamento Europeo nell'aprile del 2016 a seguito di una seconda lettura, sono due:

1. il **regolamento n. 2016/679**, entrato in vigore il 24 maggio 2016 a sostituzione della direttiva 95/46/CE e applicabile in tutti gli Stati membri a partire dal 25 maggio 2018, finalizzato a disciplinare in modo generale la protezione dei dati personali. Tale regolamento prende il nome di GDPR - *General Data Protection Regulation*, e sarà oggetto di una trattazione più approfondita nel paragrafo seguente;
2. la **direttiva 2016/680**, entrata in vigore il 5 maggio 2016, relativa al trattamento dei dati di persone fisiche da parte di autorità competenti in materia di prevenzione, contrasto e repressione dei crimini.

Entrambe le discipline citate contribuiscono alla formazione del cosiddetto *"Pacchetto europeo protezione dati"* finalizzato a fronteggiare l'aumento dei flussi di dati fra attori pubblici e privati, adeguando la *data protection* agli sviluppi tecnologici più recenti.

Ad oggi, il regolamento n. 2016/679, ancora vigente, rappresenta la principale normativa in materia di protezione dati personali.

Appare evidente come, nella sua evoluzione temporale, il diritto alla *privacy*, denominato anche *"diritto alla vita privata"* (Carnelutti, 1955), *"diritto alle vicende personali"* (Ligi, 1955) *o "diritto all'illesa intimità privata"* (Ferrara Santamaria, 1937) abbia assunto diversi significati per via della mutevolezza del contesto sociale, culturale, politico ed economico. La natura poliedrica del termine è data proprio da questa sua continua ed inarrestabile evoluzione che trova "*esplicazione in situazioni profondamente differenti che vanno dal diritto del singolo ad impedire comportamenti intrusivi nella propria vita privata ad opera dei media, al diritto di aborto, alla libertà sessuale"* (Mantelero, 2007). Negli ultimi decenni si è affermato, in opposizione al diritto di "essere lasciati soli" (Pannetta, 2006), un concetto di *privacy* inteso come *"diritto all'autodeterminazione informativa"*[7], relativo all'effettiva possibilità di

[7] Diritto all'autodeterminazione informativa di tipo costituzionale utilizzato per la prima volta nella sentenza del 15 dicembre 1983 da parte della Corte Costituzionale Tedesca.

controllare il flusso delle proprie informazioni personali. Quanto precede ha contribuito all'affermarsi della sequenza "persona-informazione-circolazione-controllo" che ha sostituito quella originaria relativa alla classica nozione di *privacy*, "persona-informazione-segretezza" (Rodotà, 1999; Westin, 1967).

Ad oggi, l'accezione più diffusa del concetto di *privacy* è proprio quella di *informational privacy*, legata al diritto del singolo di tutelarsi dalla raccolta, dall'utilizzo e/o dalla condivisione delle proprie informazioni personali. Il mondo di internet, dei *social* e delle nuove tecnologie ha fatto si che la vita privata di ogni individuo potesse essere costantemente condivisa con tutto il resto del mondo.

1.2 –PRIVACY E PROTEZIONE DATI PERSONALI: LA PERCEZIONE NEGLI USA E IN EUROPA

Ad oggi, i concetti di *privacy* e protezione dati personali sono ampiamente utilizzati e talvolta confusi in quanto ritenuti sinonimi. Essi, in realtà, sono profondamente diversi poiché provenienti da culture differenti. Infatti, mentre il modello europeo di tutela prevede la presenza di apposite autorità indipendenti, che fungano da garanti e controllino il rispetto della disciplina, il modello americano ripone grande fiducia nell'autoregolamentazione (Pagallo, 2008), non prevedendo così la presenza di specifiche autorità. In particolare, la *privacy* nasce e si sviluppa, come detto in precedenza, negli Stati Uniti attorno al 1890 con l'obiettivo di fronteggiare l'esigenza di riservatezza diffusa all'epoca tra la borghesia, proteggendo la sfera intima del singolo. Si sviluppa, quindi, come un diritto negativo con potere escludente (Rodotà, 2015), ovvero con la facoltà di privare gli altri delle proprie informazioni personali, allontanando sguardi indesiderati. La protezione dei dati personali, invece, nasce nel continente europeo alla fine degli anni '90 del secolo scorso, un secolo dopo gli USA, con la normativa 95/46/CE detta *Data Protection Directive*. La sua finalità è la tutela delle informazioni di ciascun individuo, attraverso un sistema di trattamento dei dati che assicuri la riservatezza, l'integrità e la disponibilità[8] di quest'ultimi. Per *riservatezza* si intende il grado con cui si concede l'accesso a determinate informazioni ad uno specifico gruppo di soggetti autorizzati. L'*integrità* è relativa alla presenza di informazioni aggiornate e prive di errori, caratterizzate quindi da un elevato grado di correttezza e completezza mentre la *disponibilità* riguarda il grado di accessibilità da parte degli utenti alle informazioni di cui necessitano. Appare chiaro, quindi, come la *privacy*

[8] L'articolo 32 del GDPR (regolamento 2016/679) stabilisce il cosiddetto principio RID, ossia la necessità di proteggere la riservatezza, l'integrità e la disponibilità dei dati personali.

rappresenti un diritto di tipo individuale, finalizzato a tutelare l'individuo e la sua sfera personale, mentre la protezione dei dati personali tuteli il singolo all'infuori della propria vita privata e soprattutto nelle relazioni sociali disciplinando la circolazione dei suoi dati personali. Sul web i *trend* di ricerca della parola "*privacy*" risultano sempre elevati e soprattutto nettamente superiori rispetto a quelli di "protezione dati personali", sia in Italia sia negli Stati Uniti[9]. In particolare, come è possibile notare dal grafico sottostante (Figura 1) relativo alle ricerche effettuate nell'ultimo anno (6 febbraio 2022 - 6 febbraio 2023), gli utenti hanno manifestato su base settimanale un interesse elevato, seppur altalenante, per il termine "*privacy*" mentre un interesse costantemente basso, e a tratti nullo, per la tematica "protezione dati personali".

Figura 1- Trend di ricerche "privacy"(azzurro) e "protezione dati personali" (rosso) in Italia

Fonte: Google Trends

Uno scenario analogo a quello italiano sulle ricerche effettuate nello stesso arco temporale è riscontrabile negli Stati Uniti (Figura 2). Il termine "*privacy"* gode anch'esso di una frequenza di ricerca elevata ma decrescente nel corso dell'anno, con un'inversione di tendenza in corrispondenza dell'inizio del 2023; mentre le ricerche di "*data protection"* restano sempre molto basse, ma mai nulle, a differenza di quanto avviene in Italia[10].

[9] Per semplicità, nell'analisi sulle ricerche effettuate tramite il web che verrà effettuata successivamente, negli Stati Uniti il termine "protezione dati personali" verrà considerato come "*data protection*".

[10] Grafici realizzati attraverso la piattaforma Google Trends analizzando nell'arco temporale 6 febbraio 2022 - 6 febbraio 2023 le ricerche dei termini privacy - protezione dati personali per l'Italia e privacy - data protection per gli Stati Uniti.

Figura 2 - Trend di ricerche "privacy" (azzurro) e "data protection" (rosso) in USA

Fonte: Google Trends

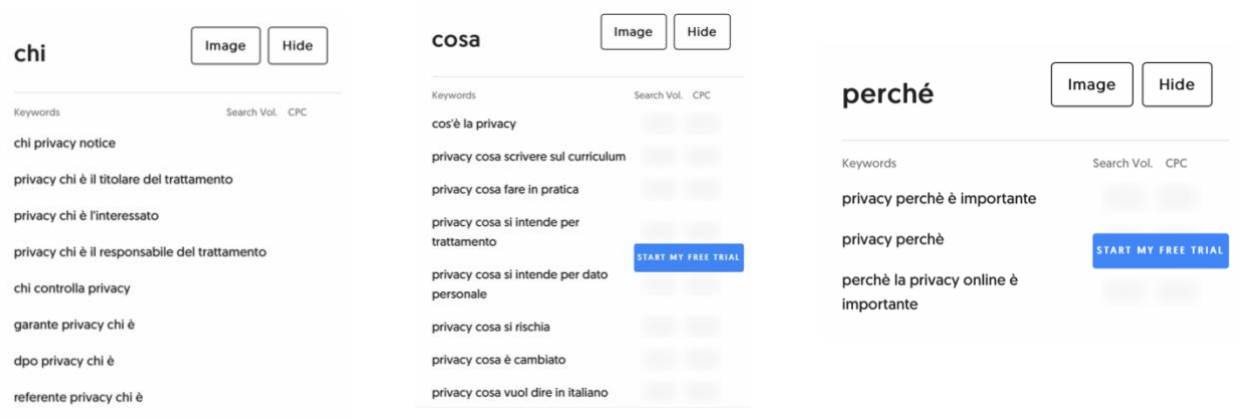
Attraverso l'analisi effettuata sui grafici precedenti emerge la presenza di un'ambiguità comunemente diffusa tra i termini "*privacy"* e "protezione dati personali/*data protection",* riscontrabile nel numero più elevato di ricerche per il primo termine, nonostante le due tematiche siano ugualmente importanti. Tale ambiguità è alimentata anche dal fatto che il termine *privacy*, al giorno d'oggi, viene utilizzato non solo in riferimento alla riservatezza del singolo, ma anche alla tutela dei dati delle persone fisiche. A tal proposito, al fine di evitare ulteriore confusione, l'autorità amministrativa indipendente, nata nel 1996 come Garante per la protezione dei dati personali[11], ha preferito mantenere su Internet l'appellativo di Garante per la Privacy, così da essere indicizzato meglio sui motori di ricerca.

Un secondo elemento utile per dimostrare la presenza di poca informazione e consapevolezza relativamente alle tematiche in questione è l'analisi delle ricerche degli utenti riferite ad una parola chiave attraverso lo strumento *'Answer The Public'*[12]. Prendendo come parola chiave "*privacy*" e considerando le principali ricerche effettuate in Italia legate al tema, è possibile notare la presenza di numerosi dubbi e questioni di carattere generale, soprattutto in risposta a tre interrogativi specifici: chi, cosa, perché (Figura 3). Ripetendo lo stesso procedimento con "protezione dati personali" come parola chiave, la maggior parte delle ricerche, sempre generiche, è riferita a tre interrogativi differenti: quale, quando, dove.

[11] Il Garante per la protezione dei dati personali nasce nel 1996 con la legge n. 675, la cosiddetta legge sulla privacy, con l'obiettivo di tutelare i diritti e le libertà fondamentali e garantire il rispetto della dignità nel trattamento dei dati personali.
A partire dal 2016, il Garante ricopre anche il ruolo di autorità designata all'attuazione del Regolamento generale sulla protezione dei dati personali (UE) 2016/679. Ad oggi, l'attività del Garante è disciplinata dal codice in materia di protezione dei dati personali (d.lg. 30 Giugno 2003 n. 196).

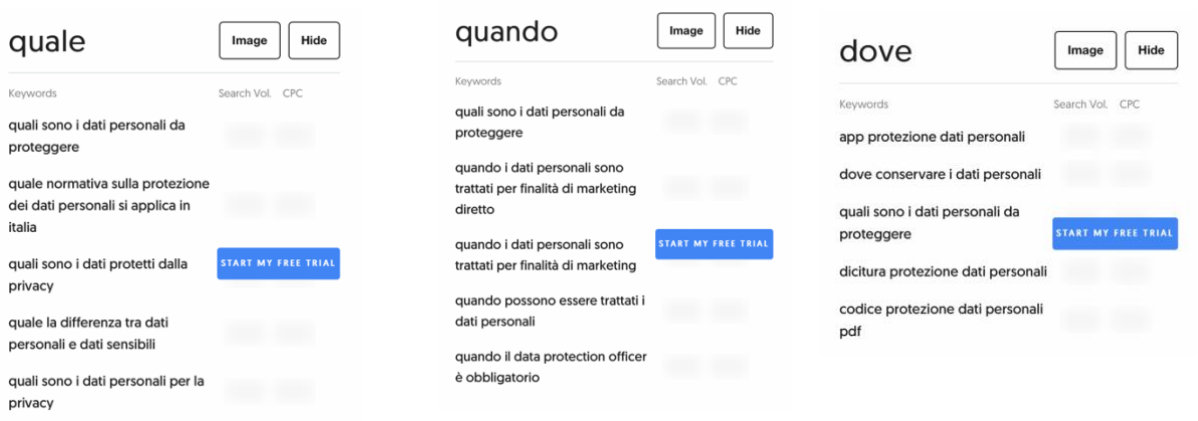
[12] Strumento di Search Engine Optimization (SEO) e di marketing usato per visualizzare sottoforma di grafici circolari e tabelle le keyword correlate ad una parola chiave e le sue previsioni di ricerca (www.answerthepublic.com).

Figura 3 - Principali ricerche relative al tema della privacy in Italia



Fonte: Answer The Public

Figura 4 - Principali ricerche relative al tema della protezione dati personali in Italia



Fonte: Answer The Public

Analogamente alle ricerche effettuate in Italia, quelle statunitensi sono tutte di carattere generale seppur in volumi molto superiori: il volume delle ricerche di "*privacy*" è 27.100 in Italia e 74.000 negli USA, mentre le ricerche di "protezione dati personali/*data protection*" si aggirano attorno ai 480 in Italia e 2.900 in USA.

Quest'analisi ci porta a riflettere su due aspetti principali. Il primo è legato alla tipologia di ricerche effettuate le quali, essendo prevalentemente non specifiche, dimostrano una scarsa conoscenza in materia e la conseguente necessità di informarsi partendo dalle basi. Tale aspetto risulta preoccupante se si considera che viviamo in una società digitale, costantemente connessa ed esposta a potenziali violazioni delle informazioni personali: è come giocare ad un gioco da tavolo senza conoscerne le regole. Il secondo aspetto è legato alla presenza di maggiori volumi

di ricerca negli Stati Uniti rispetto all'Italia, riconducibili ad avvenimenti storico-culturali che hanno influenzato profondamente la società. Gli Stati Uniti, infatti, a partire dall'attacco terroristico dell'11 settembre 2001 e passando per le rivelazioni fatte nel 2013 dall'ex dipendente della CIA e NSA Edward Snowden, hanno vissuto un crescendo di situazioni che hanno reso consapevole il popolo americano del sistema di sorveglianza e raccolta dati messo a punto dal governo nel corso degli anni. In Europa, invece, la situazione è ben differente poiché nel corso degli anni sono state adottate una serie di tutele, non presenti negli Stati Uniti, sia a livello comunitario sia a livello dei singoli Stati[13], che hanno contribuito a creare sia una maggiore consapevolezza nei cittadini, sia un sistema di tutela e protezione della *privacy* maggiormente efficace.

In sintesi, *privacy* e protezione dei dati personali rappresentano due tematiche ampiamente diffuse, ma pienamente confuse, stante la poca consapevolezza in merito alle differenze ed alle applicazioni, che assumeranno, peraltro, un peso sempre maggiore alla luce degli sviluppi tecnologici futuri. Al riguardo, una maggiore conoscenza della materia in questione è fondamentale per gestire più consapevolmente la propria vita e la propria identità in quanto, come disse Stefano Rodotà *"noi siamo i nostri dati"* (Rodotà, 2001), implicando che un utilizzo illecito di quest'ultimi rappresenti una violazione del diritto fondamentale della persona.

1.3 – *GENERAL DATA PROTECTION REGULATION* (GDPR)

Il 25 maggio 2018 ha rappresentato una tappa importante per la tutela dei dati personali a livello europeo, perché ha visto l'entrata in vigore, a pieno titolo, del regolamento 2016/679 relativo al trattamento e alla libera circolazione dei dati personali di soggetti fisici. Denominato anche GDPR, venne formulato in abrogazione della dottrina 95/46/CE, con l'obiettivo di fronteggiare le crescenti esigenze di tutela dei cittadini europei derivanti dalla progressiva diffusione di sistemi informatici che comportavano la raccolta e l'utilizzo non dichiarato di numerosi dati personali. In particolare, ha favorito l'evoluzione del concetto di *privacy*, dalla riservatezza delle proprie informazioni, alla protezione dei dati personali. Il GDPR è stato introdotto dall'Unione Europea come unico regolamento applicabile a tutti gli Stati membri, così da creare un regime giuridico omogeneo, pur consentendo ad ognuno di essi la libertà di precisarne le norme attraverso specifiche leggi nazionali[14]. Appare opportuno specificare che il GDPR tutela

[13] A livello comunitario, l'intervento principale in materia riguarda il cosiddetto GDPR, il regolamento UE 2016/679, in vigore ancora oggi mentre in Italia è rilevante la presenza del Garante per la protezione dei dati personali

[14] In Italia, per facilitare l'accettazione delle norme del GDPR, è stato emanato il d.lgs. n.101/2018 attraverso il quale si è contribuito a modificare sensibilmente la normativa italiana sulla privacy.

solo il trattamento dei dati da parte di aziende interne all'UE, ed aziende esterne all'UE che operino con i dati dei cittadini europei. A tale proposito, il regolamento vieta il trasferimento dei dati personali in Paesi al di fuori dell'Unione ed a tutte le organizzazioni internazionali che non applichino gli stessi standard di sicurezza.

1.3.1 – LE NOVITÀ INTRODOTTE DAL GDPR

Le novità apportate dal GDPR riguardano l'introduzione di norme più precise per l'informativa e il consenso, nuove sanzioni per violazioni delle norme in materia, l'imposizione di criteri più stringenti alla circolazione dei dati al di fuori dell'Unione Europea e di limiti al trattamento automatizzato dei dati. Nello specifico, uno dei principali elementi di novità è il principio di responsabilizzazione (artt. 23-25), o *accountability*, che impone ai titolari del trattamento di dati personali di adottare comportamenti proattivi, così da minimizzare i rischi per i diretti interessati e di rendere sempre conto del proprio operato. Tale principio rappresenta un importante passo in avanti in termini di protezione dei dati, poiché prevede il diretto affidamento ai titolari della facoltà di decidere le modalità, le garanzie e i limiti del trattamento dei dati, purché siano rispettati specifici criteri. L'*accountability* si fonda su due principali criteri, *privacy by design* e *privacy by default*, i quali garantiscono la tutela dei dati fin dalle fasi iniziali di ideazione e progettazione, attraverso l'adozione di strumenti e comportamenti che prevengano potenziali rischi. Al riguardo, è previsto che le aziende effettuino delle valutazioni dei rischi inerenti alle loro attività per definire le responsabilità del trattamento dei dati in capo al titolare. Inoltre, il principio stabilisce che vengano raccolti ed utilizzati solo i dati necessari per assolvere alle finalità previste, evitando così una raccolta eccessiva di quest'ultimi.

Un ulteriore elemento di novità riguarda l'ampliamento dei diritti dell'interessato con l'introduzione, fra i tanti, del diritto alla portabilità (art. 20) e del diritto all'oblio (art. 17). Il primo consente agli interessati di ricevere i propri dati e trasmetterli da un titolare del trattamento ad un altro. In tal caso, infatti, la portabilità favorisce gli utenti conferendo loro libertà di scelta e di controllo, facilitando la circolazione e il trasferimento dei dati. Il secondo conferisce ai diretti interessati la facoltà di chiedere ed ottenere la cancellazione dei propri dati. Ad oggi, il diritto all'oblio è applicabile anche online per richiedere la cancellazione dei propri dati sulla rete, nel caso in cui il loro trattamento non sia più necessario o voluto dal diretto interessato.

Un ulteriore elemento di novità introdotto dal GDPR è il *data breach* (art. 33), ossia l'obbligo per il titolare del trattamento di comunicare al garante la presenza di violazioni nei dati personali

trattati. La necessità di notifica riguarda solamente i casi in cui suddette violazioni incidano sulla disponibilità, integrità e riservatezza delle informazioni personali, con il rischio potenziale di causare danni fisici, materiali o immateriali ai diretti interessati. Nel caso in cui le imprese ed organizzazioni non rispettino gli obblighi previsti dal *data breach*, il Garante prevede sanzioni fino a 10 milioni di euro o fino al 2% del fatturato totale annuo mondiale.

Inoltre, il GDPR ha contribuito all'introduzione nel mondo del lavoro di una nuova figura professionale legata al trattamento dei dati, il *Data Protection Officer* (DPO) (art. 39). Ad esso è affidata la responsabilità di proteggere i dati personali ed assicurare la corretta gestione di quest'ultimi da parte di aziende ed enti (Soffientini, 2017). Nello specifico i compiti di un DPO sono principalmente tre:

- *informare* i titolari del trattamento in merito a come raccogliere, conservare e trattare i dati personali;
- *sorvegliare* mediante attività di controllo che le operazioni sui dati avvengono secondo le norme previste dal GDPR;
- *cooperare* fungendo da tramite tra l'autorità di controllo e il titolare del trattamento dei dati.

Fra gli elementi di maggior rilievo in materia di trattamento dei dati personali, spicca la necessaria presenza dell'informativa (o *privacy policy*) e del consenso. Secondo quanto previsto dal regolamento europeo, ogni qualvolta vi sia un trattamento di dati, il titolare del trattamento ha l'obbligo di comunicare preventivamente ai diretti interessati, attraverso l'informativa (art. 13), le finalità e modalità con cui i loro dati personali verranno utilizzati. L'informativa risponde all'esigenza di assicurare trasparenza e correttezza nei trattamenti, rendendoli legittimi attraverso il consenso del soggetto interessato. Il consenso, definito dal GDPR all'art. 4 come *"qualsiasi manifestazione di volontà libera, specifica, informata e inequivocabile dell'interessato, con la quale lo stesso manifesta il proprio assenso…"*, rappresenta una condizione fondamentale ed imprescindibile per il trattamento dei dati, revocabile in ogni caso e in qualsiasi momento. Esso deve essere libero, inequivocabile e preventivo a partire dai 16 anni di età, mentre per i minori di 16 anni[15] spetta ai genitori o a chi esercita la potestà genitoriale esprimersi a riguardo.

Attraverso il GDPR si assiste inoltre, all'evoluzione del ruolo dell'Autorità rispetto a quanto previsto dalla direttiva 95/46/CE. Secondo la direttiva, infatti, l'Autorità ricopriva un ruolo di

[15] In Italia, l'età soglia per esprimere il consenso al trattamento dei dati personali di un minore è fissata a 14 anni dall' art. 2 quinquies del d.lgs. 193/2003.

garanzia a tutela dei diritti dell'interessato, assicurandosi che quest'ultimi fossero rispettati e difesi. Nel regolamento 2016/679, invece, l'Autorità è chiamata a garantire la legittimità e il corretto utilizzo dei dati, monitorando continuamente le numerose innovazioni per evitare che possano comportare rischi. Nello specifico, l'art. 57 del GDPR evidenzia la presenza di tre fenomeni correlati tra loro che contribuiscono ad aumentare l'importanza dell'Autorità rispetto al passato (Pizzetti (a), 2016). Il primo, derivante dalla continua ed incessante evoluzione della società digitale, così come dal crescente utilizzo dei dati personali, delinea la necessità di autorità competenti e costantemente aggiornate per affrontare ogni sfida. Il secondo, riguarda l'importanza di tenere in considerazione gli effetti e le conseguenze che gli sviluppi tecnologici potrebbero avere sulle singole realtà nazionali. È opportuno, infatti, ricordare che, sebbene la società digitale sia fondamentalmente globalizzata, esistono ancora dimensioni nazionali e territoriali differenti; spetta all'Autorità garantire il rispetto delle differenze, evitando che quest'ultime creino divisioni tra Paesi. Infine, il terzo fenomeno è legato alla consapevolezza e all'importanza che la protezione dei dati personali non sia una prerogativa solo del singolo individuo, ma un'esigenza propria dell'intera società e quindi interesse pubblico. L'art. 57, inoltre, sancisce che ogni Autorità *"sorveglia gli sviluppi che presentano un interesse, se e in quanto incidenti sulla protezione dei dati personali, in particolare l'evoluzione delle tecnologie dell'informazione e della comunicazione e le prassi commerciali"*. Suddetto compito viene affidato su base generale e preventiva, per evitare che determinati sviluppi tecnologici possano incidere ed influire sulla protezione dei dati personali.

Il GDPR, quindi, sostituendo la direttiva 95/46/CE, finalizzata ad armonizzare leggi nazionali differenti, ha dato vita ad un vero e proprio salto di qualità verso una tutela di carattere più uniforme a livello europeo. Esso infatti, a differenza della direttiva sopra citata, gode di un'elevata flessibilità, in quanto gli Stati membri hanno la possibilità di dettare disposizioni integrative per facilitare l'inserimento delle norme nell'ordinamento nazionale (Pizzetti (b), 2016).

1.3.2 – DATI PERSONALI

Al giorno d'oggi, l'utilizzo massiccio ed incessante di dispositivi elettronici in ogni ambito della vita quotidiana, contribuisce a determinare una scia di tracce che ogni individuo lascia dietro di sé, usata poi per comporre ed aggiornare progressivamente l'identità digitale di ciascuno. Ogni azione compiuta che sia un *like* su Instagram, una ricerca su Google o un acquisto su Amazon, corrisponde alla creazione di enormi quantità di dati immagazzinati per poi essere utilizzati a scopi commerciali e non solo. All'interno di essi, raccolti su base

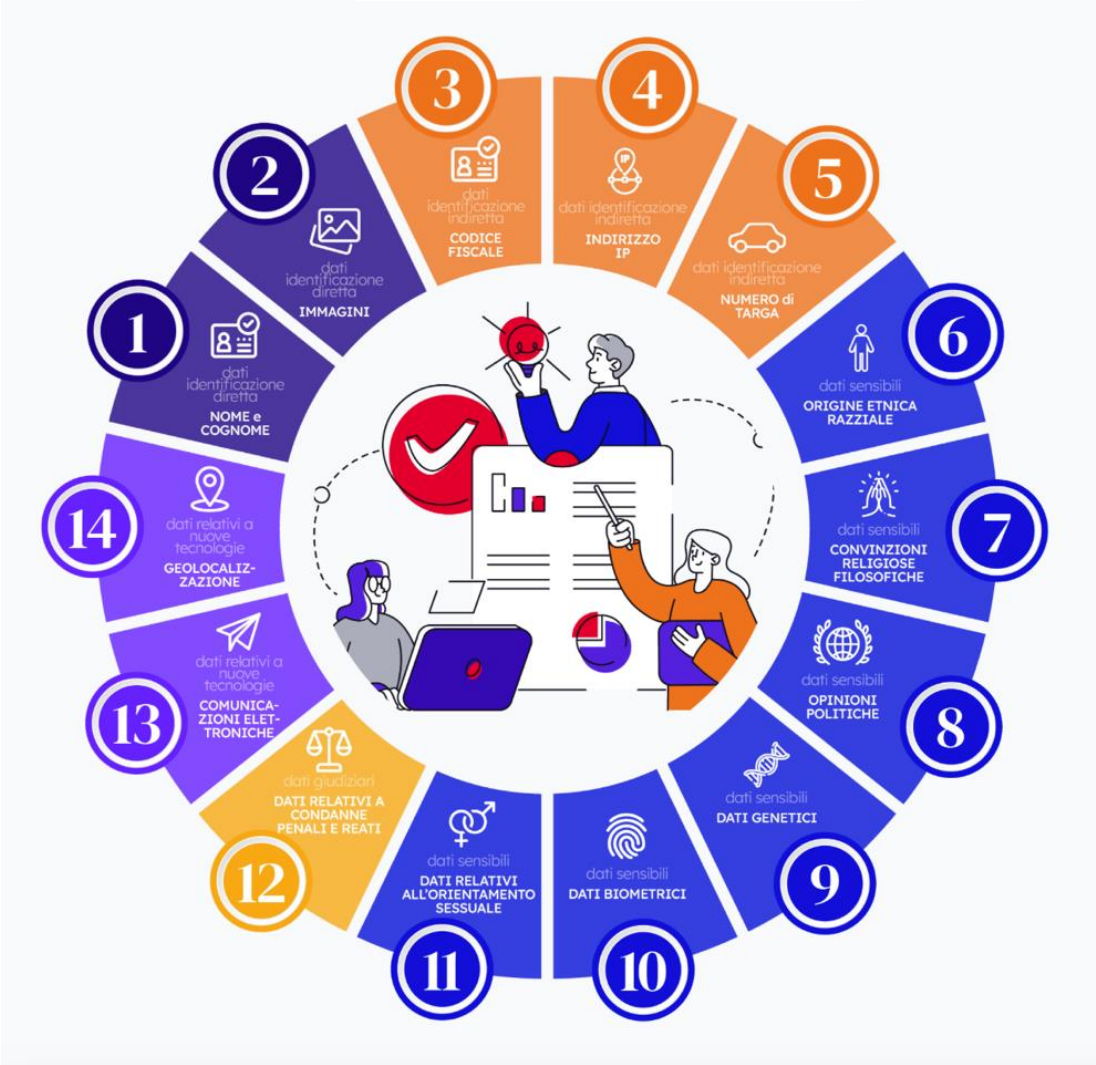
quotidiana, è possibile differenziare diverse categorie di dati a seconda del loro peso, origine e finalità di riutilizzo, tra cui i dati personali. Quest'ultimi, definiti dall'art. 4 del GDPR come *"qualsiasi informazione riguardante una persona fisica identificata o identificabile…"*, sono relativi alla sfera più intima di un individuo[16], in quanto contribuiscono a renderlo immediatamente riconoscibile e, per tale motivo, necessitano di essere tutelati. La nozione di 'dato personale' ha la caratteristica di essere elastica e funzionale a garantire una tutela adeguata a tutti (Tempestini, D'Acquisto, 2016). Rientrano, infatti, nella categoria di dati personali tutti gli elementi che, combinati tra loro, consentono di individuare l'identità di una persona e, nello specifico, non solo quelli provenienti direttamente dai soggetti interessati, ma anche quelli derivanti da soggetti terzi, differenti dal titolare del trattamento e dall'interessato, purché relativi alla sfera intima di quest'ultimo. Inoltre, vivere in una società digitale implica tenere in considerazione che le informazioni personali potrebbero derivare dall'aggregazione di vari elementi provenienti da fonti differenti senza apparenti connessioni (Buttarelli, 1997).

Diverse sono le tipologie di dati personali (Figura 5) (Welcome Digital, 2021):

- i dati anagrafici e le immagini, che garantiscono un'identificazione diretta (1-2);
- il codice fiscale, l'indirizzo IP e la targa, che garantiscono un'identificazione indiretta (3-4-5);
- l'origine etnica e razziale, le convinzioni religiose e filosofiche, le opinioni politiche, i dati genetici e biometrici e l'orientamento sessuale che costituiscono i cosiddetti dati sensibili (6-11);
- i reati e le condanne penali, dati di tipo giudiziario (12);
- la geolocalizzazione e le varie comunicazioni elettroniche, relativi alle nuove tecnologie (13-14).

[16] Il regolamento 2016/679 in materia di protezione dei dati personali esclude dal suo ambito di applicazione il trattamento dei dati personali di persone giuridiche, rimandando la tutela di quest'ultimi ai singoli Stati membri.

Figura 5 - Differenti tipologie di dati personali



Fonte: WELCOME DIGITAL

Nello specifico, i citati dati sensibili rappresentano una particolare sotto categoria dei dati personali che, secondo l'art. 9 del GDPR[17] non devono essere assolutamente trattati se non in presenza di un esplicito consenso da parte dell'interessato o per assolvere specifici obblighi.
Numerose sono le modalità utilizzate per raccogliere i dati, ciascuna delle quali restituisce un output differente in base alla loro origine. In particolare:

- i *raw data* sono dati raccolti singolarmente, non rielaborati e per tale motivo difficili da leggere ed interpretare;

[17] Con l'entrata in vigore del GDPR I dati che precedentemente appartenevano alla categoria dei dati sensibili sono stati rinominati come *dati particolari* attraverso l'art. 9, relativo al trattamento di categorie particolari di dati personali che recita: *"È vietato trattare dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, non che trattare dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona… l'interessato ha prestato il proprio consenso esplicito al trattamento di tali dati personali per una o più finalità specifiche…".*

- gli *historical data* si ottengono a partire da elementi appartenenti al passato;
- i *real data* sono rilevati al momento, in tempo reale;
- gli *expected o previsional data* derivano da ipotesi statistiche e calcolo combinatorio.

Tutti i dati vengono organizzati successivamente in insiemi detti *dataset* sulla base di caratteristiche simili o omogenee e conservati in archivi detti *database*. Esistono dati liberamente accessibili (*open data*) e utilizzabili (*open use*), in contrapposizione a quelli con accesso limitato (*closed data*) e a quelli condivisi con altri utenti (*shared data*) (Agenda Digitale (a), 2019).

A prescindere dalle caratteristiche dei dati, il trattamento di quest'ultimi è tenuto a rispettare i principi di necessità e minimizzazione. I sistemi informativi ed i vari software dovrebbero essere progettati proprio per minimizzare la raccolta dei dati, favorendone l'utilizzo solo se indispensabili per il raggiungimento delle finalità consentite. I citati due principi vietano, inoltre, il trattamento dei dati personali ove si possa raggiungere l'obiettivo da conseguire utilizzando dati anonimi, ovvero mediante modalità che consentono l'identificazione dell'interessato solo in situazioni di necessità. Un ulteriore aspetto, definito sempre dall'art. 5 del GDPR, è la liceità di trattamento dei dati personali nei confronti del soggetto interessato infatti, un trattamento è considerato lecito quando è necessario, quando avviene con il consenso dell'interessato e quando è conforme alla normativa generale.

Nelle operazioni di trattamento dei dati le parti coinvolte sono tre (Garante della Privacy): l'Interessato, il Titolare del trattamento e il Responsabile del trattamento. L'Interessato è il soggetto fisico a cui si riferiscono i dati personali; il Titolare può essere un soggetto fisico, un'autorità pubblica, un'impresa, un ente pubblico o privato che definisce le modalità del trattamento; il Responsabile, identificato anche come DPO, è il soggetto fisico o giuridico in nome del quale il titolare esegue operazioni sui dati, quindi colui che supervisiona il trattamento. Il titolare del trattamento può fare ricorso a due tecniche specifiche al fine di proteggere i dati personali dell'interessato durante le operazioni, impedendone l'identificazione. La prima è l'anonimizzazione, un processo irreversibile usato per impedire l'attribuzione di dati personali ad una specifica persona fisica, trasformandoli in dati anonimi. Il vantaggio di questa tecnica risiede nel fatto che dati anonimi non rientrano nell'ambito di applicazione del GDPR per cui sono liberamente utilizzabili, senza la necessità di rispettare gli obblighi e le responsabilità imposte dalla normativa. La seconda tecnica, la pseudonimizzazione, si differenzia dalla precedente poiché è un processo di carattere reversibile in grado di ridurre l'intellegibilità di un insieme di dati, rendendo comunque identificabile il soggetto interessato. Il GDPR, infatti, definisce all'art. 4 la pseudonimizzazione

come "*il trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive…"*, sottolineando che dati pseudonimi rientrano ancora nella categoria dei dati personali, poiché attribuibili al diretto interessato attraverso l'utilizzo di informazioni aggiuntive. La crittografia, ad esempio, può rappresentare uno strumento di pseudonimizzazione, in quanto utilizzando chiavi crittografiche trasforma le informazioni in messaggi difficilmente comprensibili, consentendo un buon grado di riservatezza. Il messaggio originale risulta comunque accessibile, ma solo a seguito di un processo di decrittazione che rende possibile la leggibilità dei dati personali in esso contenuti.

Per quanto precede, quindi, è possibile affermare che l'anonimizzazione si pone a tutela della *privacy*, evitando che un dato possa essere attribuito direttamente ad una persona (Tempestini, D'Acquisto, 2016), mentre la pseudonimizzazione rappresenta uno strumento a tutela della sicurezza dei dati personali, volto a garantirne la confidenzialità e l'integrità.

Il regolamento 2016/679 ha rappresentato e rappresenta ancora oggi il principale intervento giurisprudenziale, a livello europeo, in materia di protezione dei dati personali. Esso infatti, si pone come una normativa solida ed al contempo flessibile, adatta a fronteggiare tutti i possibili sviluppi futuri della società digitale. Il GDPR ha favorito, inoltre, l'affermarsi della consapevolezza ormai diffusa secondo la quale la tutela dei dati personali non è più solo una prerogativa che spetta al singolo ma è una garanzia offerta alla società intera. A tale proposito, le autorità sono chiamate a ricoprire un ruolo sempre più importante e centrale, garantendo legittimità nel trattamento dei dati e tutelando le libertà dei cittadini.

Nella società odierna, come visto in precedenza, l'utilizzo massiccio dei dispositivi digitali sia in ambito pubblico, sia privato, ha reso gli individui sempre più connessi e trasparenti, favorendo la creazione di enormi quantità di dati personali quale conseguenza di una raccolta sempre più pervasiva da parte di organizzazioni private e pubbliche. Il GDPR, quindi, nasce dall'esigenza di tutelare la sfera più intima di ogni individuo, che oggi si estrinseca nella sua identità digitale.

1.4 – PRIVACY E PROTEZIONE DATI PERSONALI NELLA SOCIETÀ DIGITALE

La società digitale in cui viviamo oggi è frutto di un'evoluzione di lungo periodo, iniziata già nel secolo scorso e che ha coinvolto ogni ambito della vita quotidiana, da quello sociale a quello economico, politico e culturale. L'introduzione del *personal computer* nella quotidianità delle persone, a partire dalla metà degli anni '80, ha rappresentato il primo vero passo in avanti che,

unito al successivo sviluppo di reti digitali e mezzi di telecomunicazioni, ha innescato tale processo di trasformazione. Alla società digitale si deve il merito di aver soppiantato quella industriale, diffondendo una comunicazione legata prevalentemente all'utilizzo di Internet, di strumenti digitali e reti sociali. A tale proposito, secondo il sociologo Manuel Castells, la nascita dei *social media* ha favorito l'affermarsi della comunicazione di massa del sé (*mass-self-communication*): una nuova forma di comunicazione, attraverso la quale raccontarsi agli altri ed esprimersi, fondamentale per la costruzione dell'identità digitale di ciascuno (Castells, 2009).

Lo sviluppo tecnologico, la disponibilità di sistemi informatici e la versatilità della rete Internet hanno contribuito a facilitare la vita ed il lavoro degli individui. Tuttavia, l'impiego massivo di tale tecnologia ha parimenti comportato la creazione di dati personali che, a causa della loro mole, diventano sempre più difficili da proteggere tutelando gli stessi cittadini. Tali dispositivi e strumenti appartengono al cosiddetto *Internet of Things*.

1.4.1 – *INTERNET OF THINGS*: IL FENOMENO DEGLI OGGETTI "ATTIVI"[18]

In un'epoca di grandi trasformazioni e cambiamenti, come quella in cui si è diffusa la società digitale, anche gli oggetti hanno assunto un ruolo fondamentale. L'espressione *Internet of Things* (IoT), infatti, è un neologismo[19] usato nel mondo delle telecomunicazioni per indicare un insieme di tecnologie in grado di collegare ad Internet qualunque oggetto o dispositivo. Inizialmente, tale fenomeno riguardava smartphone e computer per poi estendersi, in un secondo momento, anche alle numerose tecnologie indossabili (cosiddette *wearables*) e ai dispositivi *smart* per la casa (Di Landro, 2020). Ulteriori ambiti di applicazione dell'IoT includono i veicoli auto pilotati, le operazioni chirurgiche effettuate mediante robot, i processi industriali ed aziendali, i sistemi di videosorveglianza a distanza tramite *smartphone* e numerosi altri ancora, a testimonianza di come gli *smart objects* abbiano profondamente rivoluzionato il mondo attuale.

La prospettiva di un futuro di comodità, *smart* ed iper connesso è allettante, ma porta con sé elevate complessità e difficoltà di gestione. Tutti gli oggetti connessi alla rete, infatti, rappresentano degli strumenti per la raccolta e l'acquisizione di dati personali. È sufficiente pensare alla quantità di dispositivi *smart* che ognuno possiede per capire quanto il fenomeno

[18] A.C. DI LANDRO, *Big Data. Rischi e tutele nel trattamento dei dati personali*, Edizioni Scientifiche Italiane, 2020, p.40.

[19] Il neologismo IoT è stato usato per la prima volta nel 1999 da Kevin Ashton, ricercatore del MIT, nel tentativo di dare un nome agli oggetti reali connessi ad internet.

dell'IoT possa essere rischioso sia per la sicurezza, sia per la tutela della *privacy*. In termini di sicurezza, tutti gli strumenti che funzionano tramite Internet sono dei potenziali punti di accesso per la realizzazione di *cyber* attacchi, finalizzati ad acquisire dati personali, a riprogrammare i dispositivi, a causare malfunzionamenti o a estorcere denaro. Inoltre, vivere in un mondo iper connesso implica lo svolgimento delle principali attività quotidiane attraverso sistemi connessi ad Internet, per cui la crescente diffusione di oggetti IoT contribuisce a rendere sempre più impensabile e complessa una vita senza connessioni in rete (Internet Society, 2015). Un esempio di quanto sopra citato deriva dai frequenti episodi di '*Instagram down'* o '*WhatsApp down'* che coinvolgono migliaia di utenti, talvolta per svariate ore, creando pesanti ripercussioni sia a livello sociale, sia lavorativo. Infatti, in tali situazioni, è come se le persone fossero paralizzate ed incapaci di comunicare se non tramite suddette piattaforme, nonostante esistano numerose altre soluzioni, dimostrando quanto Internet abbia rivoluzionato la vita dell'uomo, rendendolo tanto libero quanto schiavo.[20]

[20] All'interno dell'espressione 'Libero quanto schiavo' l'uomo viene considerate libero in quanto, l'introduzione di Internet e delle tecnologie digitali, ha facilitato la sua quotidianità, le comunicazioni, gli spostamenti ed ha inoltre contribuito a migliorare la qualità della sua vita in molti ambiti soprattutto quello medico e quello lavorativo. Quanto sopra citato, però, rappresenta solamente una faccia della medaglia poiché l'uomo è libero, ma al contempo anche schiavo di tutti i moderni meccanismi digitali, senza la possibilità di sottrarsi ad essi, in quanto sono permeati in ogni ambito e senza di essi non sarebbe possibile effettuare la maggior parte delle azioni quotidiane.

Tabella 1 - Ambiti di applicazione dell'IoT

Ambiti di applicazione dell'IoT[21]	
AMBITO	**ESEMPIO**
Esseri umani	Dispositivi per monitorare il benessere e la salute, per gestire patologie (es. pacemaker)
Casa	Dispositivi di sorveglianza ed elettrodomestici tecnologici controllabili a distanza
Luoghi pubblici	Negozi, ristoranti, banche e in generale ogni posto in cui i consumatori/clienti possono acquistare prodotti o servizi
Ufficio	Dispositivi per garantire sicurezza ed aumentare la produttività dei dipendenti
Imprese ed aziende	Luoghi con routine lavorativa e ripetitive come ospedali e fabbriche in cui è importante e necessario incrementare sempre l'efficienza e la produttività
Veicoli	Veicoli quali auto e motoveicoli, camion, natanti, aerei e treni che necessitano di dispositivi elettronici per il funzionamento e il mantenimento, per statistiche di funzionamento, sistemi di pilotaggio automatico
Città	Dispositivi per gestire spazi pubblici ed infrastrutture, controllo del traffico e monitoraggio ambientale

Fonte: McKinsey Global Institute

Come accennato in precedenza, l'*Internet of Things* comporta dei rischi anche in termini di tutela della *privacy* per via delle grandi quantità di dati create e raccolte. In merito, il problema principale risiede nell'aggregazione di dati provenienti da diversi dispositivi e conseguentemente nella vendita di essi a terze parti. Così facendo, infatti, è possibile profilare ciascun individuo, delineandone un quadro più preciso. Per evitare quindi che gli utenti si sentano minacciati durante l'utilizzo di Internet, perdendo fiducia, è fondamentale dichiarare in modo aperto e trasparente le finalità di raccolta ed utilizzo dei dati. Esistono, inoltre, numerosi accorgimenti a cui prestare attenzione, soprattutto in vista dei futuri sviluppi tecnologici. In

[21] M. JAMES, M. CHUI, P. BISSON, J. WOETZEL, R. DOBBS, J. BUGHIN, D. AHARON, *Mapping the world beyond the hype*, McKinsey Global Institute, 2015.

primis, è fondamentale gestire in modo accurato la sicurezza delle proprie reti e connessioni, per evitare intrusioni ed attacchi da parte di *hacker*. In aggiunta, per tutelarsi al meglio, è consigliabile proteggere i propri accessi tramite un'autenticazione a due fattori ed effettuare sempre comunicazioni crittografate.

Attorno al fenomeno dell'*Internet of Things*, soprattutto nell'ultimo decennio, alla luce dei numerosi scandali in materia di *privacy* e protezione dati personali, è venuto a crearsi un acceso dibattito tra due visioni contrapposte. Da un lato si trovano tutti coloro che considerano l'IoT rivoluzionario e in grado di stimolare il progresso, generando una maggiore efficienza e creando numerose opportunità a livello globale; dall'altro lato, invece, si posizionano tutti quei soggetti che considerano tale fenomeno in modo negativo, perché lo ritengono un enorme tentativo di sorveglianza di massa messo appunto tramite violazioni a livello di *privacy* e sicurezza (Internet Society, 2015).

In ogni caso, per quanto l'*Internet of Things* divida l'opinione pubblica, esso rappresenta comunque una forma di evoluzione che coinvolge l'uomo nel suo rapporto con l'uso di Internet, e per questo si ritiene che in futuro, vista la pervasività degli *smart object*s, il valore economico generato da esso sarà molto elevato e continuamente in crescita. Nello specifico, secondo il McKinsey Global Institute, il valore economico potenziale a livello globale generato dall'IoT entro il 2030 potrebbe variare dai 5.5 ai 12.6 trilioni[22] di dollari, provenendo maggiormente dal settore industriale, per il 26% ($1.4 - $3.3 trilioni), e dal settore della salute umana, per il 10-14% ($0.55 - $1.7 trilioni) (Chui, Collins, Patel, 2021). Se tali previsioni si rivelassero corrette, in futuro si assisterebbe alla crescita di nuove numerose opportunità e sfide globali.

1.4.2 – RISCHI POTENZIALI PER GLI UTENTI DI UNA SOCIETÀ DIGITALE

"Le tecnologie digitali portano enormi benefici alla società, ma la sorveglianza pervasiva a un costo elevato, perché mina i diritti e soffoca lo sviluppo di democrazie vibranti e pluralistiche".

Queste le parole che l'Alto Commissario per i Diritti Umani, Mrs. Nada Al-Nashif, ha scelto di utilizzare per sottolineare l'importanza di tutelare la propria *privacy* e i propri dati dalle numerose minacce provenienti dall'esterno. La nascita di Internet, infatti, ha da un lato favorito la diffusione di un forte entusiasmo tra gli individui che lo consideravano come una "nuova dimensione" attraverso la quale poter effettuare azioni impensabili nella vita reale, ad esempio: comunicare istantaneamente con amici dall'altra parte del globo, effettuare rapide ricerche bibliografiche da casa senza il bisogno di recarsi in biblioteca e molto altro ancora. Dall'altro

[22] Un trilione corrisponde ad un milione di bilioni (10^{18}).

lato, invece, essa ha comportato la creazione di innumerevoli rischi legati alle grandi quantità di dati raccolti su base quotidiana ed utilizzati per definire l'identità *social* di ciascuno. Tale identità rappresenta un elemento talvolta complesso, poiché si costituisce a partire dalle numerose interazioni online tra individui, dalle loro pratiche di *networking*, quando diventano virtualmente "amici", quando ricevono dei *like*, quando inseriscono dei commenti o quando compiono qualsiasi altro tipo di azione rilevante. Uno dei potenziali rischi in cui gli utenti possono imbattersi e la possibilità di perdere il controllo della propria identità virtuale, in quanto sui *social network* è difficile controllare i contesti in cui le proprie informazioni vengono condivise; quest'ultime, infatti, possono essere viste da chiunque ed utilizzate, addirittura, da sistemi di intelligenza artificiale per effettuare analisi e valutazioni. La perdita di controllo della propria identità può avvenire anche per mano di un *cyber*criminale, il quale agisce attraverso l'accesso illegale e non autorizzato ai dati personali di un altro soggetto, effettuando un furto di identità digitale. Il reato sopra citato, per via della grande diffusione di dispositivi e tecnologie digitali, è sempre più frequente anche nel nostro paese, che si posiziona 12º nella classifica internazionale relativa alle violazioni degli account.[23] Le modalità attraverso le quali operano i criminali digitali sono numerose tra cui le principali sono:

- il ***PHISHING***: tecnica che prevede l'invio di messaggi o *e-mail* molto simili a comunicazioni ufficiali, in cui è richiesto alla vittima l'inserimento di informazioni personali;
- i ***MALWARE***: tecnica che prevede l'utilizzo di virus o *keylogger*[24] per intercettare ed immagazzinare le *password* degli utenti. I *malware* accedono ai dispositivi infettati tramite Internet e posta elettronica, agendo segretamente senza che gli utenti ne siano consapevoli;
- gli ***SPYWARE***: tipologie di *malware* difficilmente rilevabili, usate per raccogliere dati di navigazione, cronologia delle ricerche, informazioni personali come la carta di credito e molto altro ancora. Tipicamente gli *spyware* provengono da altri software, da file scaricati gratuitamente come musica o film o da allegati via mail.

Tali modalità appartengono tutte alla sfera del *social engineering*, ossia un insieme di tecniche che non si configurano come un attacco diretto ai danni delle vittime ma piuttosto come un'operazione di psicologia inversa. Essa, infatti, consiste nell'ottenere dati personali direttamente dai soggetti interessati, spingendoli a rivelare informazioni o a consentire

[23] La classifica proviene dall'analisi effettuata nel 2021 da CybergON, business unit di Elmec informatica operante nel campo della cybersecurity.

[24] I *keylogger* sono degli strumenti di hardware o software utilizzati per registrare ed immagazzinare tutto quello che viene digitato sulla tastiera di un computer senza che l'utente se ne accorga.

l'accesso a specifici sistemi informatici. Le conseguenze per chi subisce un furto di identità possono avere differenti livelli di gravità e sono relative sia ad aspetti finanziari, sia legali. Dal punto di vista finanziario, c'è il rischio che i criminali sfruttino l'accesso a carte di credito e conti bancari per impadronirsi di tutte le risorse della vittima, ovvero effettuare costosi acquisti ed operazioni. Dal punto di vista legale, invece, è possibile che la vittima si trovi ad essere responsabile di atti illeciti compiuti in suo nome. A tal proposito, è importante assumere sempre una serie di precauzioni per prevenire il furto della propria identità digitale: le *password* a protezione degli *account* dovrebbero essere differenziate per ogni piattaforma, complesse, non scontate e cambiate periodicamente. Inoltre, è consigliabile l'utilizzo di un'autenticazione a due fattori, di *browser* e connessioni Internet sicure e il possesso di un *antivirus*.

Un ulteriore rischio deriva dalla creazione e diffusione di informazioni personali false al fine di dare vita ad identità digitali fittizie molto diverse dalla realtà dette *"catfish"*. Tale tecnica è utilizzata soprattutto per effettuare frodi e truffe ai danni degli utenti.

Infine, il rischio più temuto, soprattutto negli ultimi tempi a seguito della vertiginosa crescita degli episodi, è quello dell'attacco *hacker*. Secondo il rapporto di fine anno redatto dalla Polizia Postale Italiana, nel 2022 si è assistito ad un incremento del 138% nel numero di attacchi informatici subiti, passando da 5434 nel 2021 a 12.947 nel 2022 (Commissariato di PS, 2023). Ad esserne maggiormente colpiti sono principalmente aziende e Piccole Medie Imprese (PMI) operanti in settori strategici, ma anche enti governativi, militari e finanziari per via di quanto avviene nel contesto geopolitico attuale.[25] Le tecniche più utilizzate prevedono:

- **attacchi *Ransomware*:** finalizzati a paralizzare e bloccare l'accesso a computer, servizi e sistemi informatici attraverso la cifratura dei dati in essi contenuti. L'utilizzo della tecnica del *ransomware* è costantemente in crescita per via dell'elevata possibilità di lucro. Tali attacchi, infatti, si configurano come dei veri e propri ricatti in quanto i *cyber*criminali tendono a prendere in ostaggio dati ed informazioni fondamentali e di grande valore assicurandosi, così, che le vittime siano disposte a pagare elevate somme di denaro pur di riappropriarsi dei propri dati[26];
- **attacchi DDoS (*Distributed Denial of Service*):** finalizzati a sovraccaricare di false

[25] Il conflitto fra Russia e Ucraina si trova ad avere importanti ripercussioni non solo a livello geopolitico ma anche in termini di *cybersecurity*. In rete, infatti, numerosi *hacker* si sono schierati a favore della Russia e altrettanti dell'Ucraina portando la guerra anche nel mondo virtuale. Una guerra, quella virtuale, combattuta attraverso attacchi ostili mirati a paralizzare servizi e sistemi fondamentali.

[26] Durante il periodo di piena emergenza da COVID-19 diversi ospedali e servizi sanitari italiani sono stati i principali bersagli di numerosi attacchi *hacker* per via della crucialità ed elevate importanza dei dati dei pazienti. Tutto ciò ha sicuramente contribuito a peggiorare una situazione già critica, evidenziando limiti e falle dei sistemi di sicurezza presenti nel nostro paese.

richieste di accesso siti web e risorse di rete, rendendoli così inutilizzabili. Sebbene tali attacchi non vengano utilizzati per furti di informazioni, risultano ugualmente dannosi e costosi per i soggetti e le organizzazioni coinvolte.[27] Essi avvengono, infatti, su ampia scala e le richieste di accesso fasulle provengono contemporaneamente da più fonti. Esiste, inoltre, una tipologia di attacco dal funzionamento analogo al DDoS denominata DoS (Denial of Service) che differisce per il fatto che la fonte di traffico è unica;

- **attacchi ATP (*Advanced Persistent Threat*):** attacchi realizzati su larga scala da soggetti esperti e dotati di ingenti risorse tecnologiche ed economiche. l'approccio adottato è di tipo persistente poiché essi non sono finalizzati al raggiungimento immediato degli obiettivi, ma a mantenere l'accesso a sistemi informatici per periodi di tempo più estesi. Tali attacchi sono principalmente rivolti a grandi imprese e Stati nazionali con l'obiettivo di ottenere un vantaggio competitivo attraverso la raccolta di informazioni, segreti industriali militari, brevetti e proprietà intellettuali.

La maggior parte degli attacchi *hacker* e dei furti di identità esaminati, avvengono attraverso l'utilizzo di dati, contenuti e strumenti presenti all'interno di una parte di Internet denominata *Dark Web*. Esso rappresenta un luogo virtuale accessibile solo mediante specifici *browser* (TOR) e connessioni all'interno del quale è possibile trovare svariati contenuti, prodotti e servizi da acquistare illegalmente quali: carte di credito clonate e documenti falsi, droga, armi, password e account, malware pronti all'utilizzo, materiale pornografico e molto altro ancora. Poiché tutto quello che avviene sul *Dark Web* è illegale, è fondamentale che gli *hacker* lavorino e navighino al suo interno nel completo anonimato, così da evitare di essere intercettati dalle autorità competenti. Il tutto risulta paradossale se si pensa che al suo interno, invece, la *privacy* di moltissimi soggetti talvolta inconsapevoli viene meno per via di numerosi dati ed informazioni personali rubati e poi rivenduti tramite esso a soggetti terzi o usati come pretesto per un riscatto. Tale situazione è esemplificativa della realtà a noi contemporanea, una realtà in cui tutto ruota attorno ai dati seguendo una logica puramente materialistica, dove quest'ultimi sono una fonte di *business* da cui trarre profitto per le aziende e beneficio per gli individui, indirizzando gli acquisti e cercando di rispettare la loro sfera privata. L'odierna logica materialistica che sottende alla raccolta e all'utilizzo dei dati, si è tristemente diffusa ed affermata nel corso degli anni affiancata, paradossalmente, dalla crescente consapevolezza degli utenti sui rischi e sulle minacce provenienti dalla rete. In particolare, la consapevolezza

[27] Solo per citare un esempio, il 22 febbraio 2023, ventiquattr'ore dopo la visita della premier Meloni in Ucraina, l'Italia subisce un attacco da parte di un gruppo di *hacker* filorussi che va a colpire direttamente il sito del Ministero degli Esteri e dei Carabinieri. Tale attacco evidenzia un pericolo importante, ossia la potenziale paralisi di attività cruciali per il paese quali servizi istituzionali, bancari e infrastrutture.

di ognuno di noi è stata recentemente aggiornata da tutta una serie di conoscenze in materia, conseguenti alle rivelazioni di Edward Snowden, che hanno svelato agli occhi del mondo un utilizzo di dati personali ritenuto improprio e lesivo della *privacy*.

1.5 – LO SCANDALO DELLA SORVEGLIANZA DI MASSA: IL CASO SNOWDEN

Lo scandalo della fuga di notizie, scoppiato nel 2013 e passato alla storia con il nome di Datagate[28], non è altro che il risultato di una serie di conseguenze all'attacco terroristico dell'11 Settembre 2001, che ha colpito nella sicurezza e nell'orgoglio il popolo e le istituzioni americane. A seguito di esso, infatti, il governo americano ha incrementato vertiginosamente i controlli su Internet e sui dispositivi digitali, creando un sistema di sorveglianza di massa atto ad evitare il ripetersi di simili episodi. Il monitoraggio risponde ad esigenze di *"foreign intelligence"*[29] che comportano la raccolta e il trattamento di dati non in modo mirato (Big Data), a differenza del sistema degli *'small data'*[30], che veniva usato, in precedenza, per la raccolta di dati mirati a singole persone (Hu, 2015). La raccolta dei *Big Data* prevede l'acquisizione automatica di dati in grandi quantità, la loro conservazione per lunghi intervalli di tempo, l'integrazione con altre banche dati e, infine, l'analisi tramite specifici software informatici. La complessità e la laboriosità nell'acquisire dati ed informazioni permettono di classificare la sorveglianza del governo statunitense come *"un formidabile sistema di controllo a scala mondiale*" (Pizzetti, 2013) e contribuiscono a tracciare *"il panorama di un vero e proprio controllo […] inammissibilmente intrusivo sia della privacy delle persone coinvolte che della libertà stessa degli Stati"* (Pizzetti, 2013). Nello specifico, il *Datagate* scoppia a seguito di una serie di rivelazioni fatte da un ex dipendente della NSA[31] e collaboratore della CIA, Edward Snowden, il quale, per l'incarico ricoperto, aveva accesso a server e documenti riservati del governo statunitense. Nonostante tale azione fosse rischiosa per la sua incolumità,

[28] Il caso *Datagate* è considerato, ad oggi, uno dei principali scandali causati dalla rivelazione e divulgazione di informazioni e documenti segreti.

[29] Per *foreign intelligence* si intende un servizio che fa parte del sistema di sicurezza nazionale finalizzato a proteggere i cittadini e lo Stato da minacce provenienti dall'esterno.

[30] A differenza dei *Big Data*, che si riferiscono a grandi quantità di informazioni generiche proprie di una massa, gli *Small Data* sono dati propri di singole persone, usati per tenere in considerazione la componente umana durante il processo di redazione di un una strategia di marketing. Tali dati hanno assunto un'importanza sempre maggiore poiché consentono alle imprese di capire esattamente i bisogni e le necessità della clientela, e offrire, così, un significativo contributo al marketing emozionale.

[31] La *National Security Agency* (NSA) è l'organo statunitense, appartenente al Dipartimento della Difesa, che si occupa della sicurezza nazionale insieme alla CIA e all'FBI. Nello specifico, essa ha il compito di monitorare il territorio nazionale per proteggerlo da possibili attacchi, proteggendo inoltre i dati degli uffici governativi.

egli decise ugualmente di portare a termine quello da lui definito come *"uno sforzo per informare il pubblico su ciò che viene fatto in loro nome e quello che è fatto contro di loro"*[32]. Tutto ebbe inizio il 5 giugno 2013 con la pubblicazione esclusiva, da parte del quotidiano britannico 'The Guardian', di informazioni riguardanti un programma di sorveglianza di massa operato nei confronti di cittadini americani e stranieri. Tale programma prevedeva l'accesso a dati, informazioni personali e tabulati telefonici di migliaia di utenti grazie alla collaborazione con l'azienda di telecomunicazioni 'Verizon'. Le modalità attraverso le quali avveniva la raccolta dei dati erano due: tramite intercettazioni dirette di comunicazioni telefoniche e telematiche o accedendo ai dati degli utenti relativi al traffico (Resta, 2015). Successivamente, lo scandalo si allargò rapidamente grazie ad ulteriori dichiarazioni che sostenevano il coinvolgimento di grandi aziende del calibro di Google, Facebook e Apple. Quest'ultime rappresentano solo tre delle nove aziende coinvolte dall'NSA nell'ambito del programma denominato *Planning tool for Resource Integration Synchronization and Management* (PRISM). Il PRISM era un programma di sorveglianza, attivo già dal 2007 in virtù del *Patriot Act*[33], finalizzato a monitorare su scala mondiale le comunicazioni effettuate tramite la rete Internet. Esso venne definito dal giornalista Glenn Greenwald come il risultato di *"una moltitudine di negoziati segreti intercorsi tra la NSA e i colossi tecnologici, nel corso dei quali l'Agenzia ha esercitato pressioni per garantirsi un accesso senza restrizioni ai loro sistemi informatici"* (Greenwald, 2014). I dati provenivano direttamente dai server dei provider (The Washington Post, 2013) e riguardavano principalmente e-mail, messaggi, chiamate vocali, videochiamate, foto e video successivamente catalogati attraverso un software detto *Boundless Informant*[34]. Le nove aziende coinvolte nel programma appartenevano al settore della tecnologia e delle comunicazioni e comprendevano: Aol, Apple, Facebook, Google, Microsoft, PalTalk, Skype e Yahoo. Il fatto che la sorveglianza operata dal programma PRISM avvenisse a livello mondiale, comportava il monitoraggio di dati, attività e comunicazioni non solo di cittadini statunitensi, ma anche stranieri. Ciò suscitò molte polemiche soprattutto in Europa a

[32] Fu poi accusato dalla procura federale di spionaggio e furto di proprietà del governo. Ad oggi, Edward Snowden vive in esilio a Mosca. Nel 2022 ha giurato fedeltà alla federazione russa, ottenendo in cambio il passaporto e diventando, a tutti gli effetti, un cittadino russo. In realtà I suoi progetti erano altri. Nel 2013 infatti, dopo essere volato ad Hong Kong ed aver fatto le sue rivelazioni da lì, l'obiettivo era di spostarsi in Ecuador, paese a cui voleva chiedere asilo, ma durante lo scalo a Mosca gli venne ritirato il passaporto impedendogli di viaggiare nuovamente.

[33] Il *Patriot Act* è una legge federale statunitense che nasce nell'ottobre del 2001 a seguito dell'attentato alle Torri Gemelle, per fronteggiare le minacce subite dal terrorismo. Nello specifico, la sezione 215 conferisce alle autorità federali e agli organi di polizia l'accesso alle informazioni possedute dagli Internet Service Providers (ISP). Suddetta legge rappresenta una modifica del precedente *Foreign Intelligence Surveillance Act* (FISA) emanato nel 1978.

[34] *Boundless Informant* è lo strumento di data mining utilizzato dalla NSA per visualizzare ed analizzare Big Data.

causa dell'incompatibilità del PRISM con la normativa comunitaria[35]. L'Europa, infatti, è da sempre un continente attento e sensibile a questioni legate alla *privacy* e alla tutela dei dati personali, tanto che le tensioni con gli Stati Uniti, scaturite da tale vicenda, spinsero l'Europa a rivedere e modificare le normative in materia di trasferimento dei dati oltre oceano e a rafforzare la tutela dei dati personali.

In sintesi, quindi, il Datagate destò grande scalpore e al contempo preoccupazione tra gli individui, non tanto per il fatto che essi si sentissero costantemente sorvegliati, quanto più per la raccolta preventiva di dati ed informazioni in modo indiscriminato. Quest'ultima, infatti, avveniva senza tenere in considerazione le numerose garanzie costituzionali e norme nazionali (Ziccardi, 2014) in materia e, non a caso, ebbe importanti ripercussioni sia a livello politico, contribuendo a mutare gli equilibri internazionali, sia a livello giuridico.

[35] All'epoca la normativa comunitaria in vigore era la 95/46/CE poi sostituita dalla n. 2016/679, anche detta GDPR.

2. IL RUOLO DEI *SOCIAL MEDIA* COME FONTE DI DATI

"I social media sono una grande risorsa per l'apprendimento e la condivisione di conoscenze, ma è importante filtrare le informazioni e cercare la verità."
(Bill Gates)

2.1 – *SOCIAL MEDIA*: DEFINIZIONE E CARATTERISTICHE

Negli ultimi decenni si è assistito all'emergere ed alla diffusione di nuovi mezzi di comunicazione legati ad Internet ed alla rete: *i social media*. Essi hanno avuto un impatto rivoluzionario sulla quotidianità, assumendo un ruolo rilevante ed essenziale sia per gli individui, sia per la società. Nel corso degli anni, inoltre, quest'ultimi si sono evoluti, passando da semplici sistemi di messaggistica a veri e propri *social network*, forum di discussione, piattaforme di condivisione di contenuti multimediali e molto altro. Il continuo mutamento sotteso al fenomeno dei *social media* ha portato alla formulazione di diverse definizioni su di esso, tuttavia, quella più accettata e consolidata, ad oggi, è quella proposta da Kaplan e Haenlein (2010):

"I social media sono un insieme di applicazioni internet-based che sono costruite sulle fondamenta ideologiche e tecnologiche del Web 2.0 e che consentono la creazione e lo scambio di contenuto user-generated (UCG)." (Kaplan & Haenlein, 2010)

Secondo i due studiosi, infatti, i *social media* sono piattaforme online che consentono agli utenti di creare e condividere contenuti generati in prima persona, partecipare ad interazioni sociali e connettersi con altri membri della *community*. Tale definizione cattura l'essenza del concetto e riflette la natura dinamica ed interattiva di queste piattaforme. All'origine dell'ampia diffusione e della grande popolarità riscossa dei *social media* è possibile individuare diversi fattori chiave:

- **connessione globale:** i *social media* hanno abbattuto le barriere geografiche e facilitato la connessione tra gli individui a livello globale. Questa possibilità di comunicare e di interagire con individui provenienti da diverse parti del mondo li ha resi un potente strumento di connessione e condivisione;
- **accessibilità e facilità d'uso:** le piattaforme di *social media* sono diventate sempre più accessibili ed intuitive nel corso del tempo. Al riguardo, infatti, la maggior parte delle persone può utilizzarli con relativa facilità, anche senza competenze tecniche avanzate.

Tutto ciò ha contribuito ad una maggiore adozione e partecipazione da parte di un vasto pubblico;

- **interazione e coinvolgimento:** i *social media* offrono numerosi strumenti per l'interazione ed il coinvolgimento degli utenti. Le persone possono, infatti, lasciare commenti, mettere "mi piace", condividere e retweettare contenuti, partecipare a sondaggi, creare discussioni e persino collaborare a progetti comuni. Tale interattività ha contribuito a creare un senso di comunità e a stimolare l'engagement degli utenti;
- **opportunità di business:** i *social media* offrono alle aziende e agli operatori di marketing un canale per raggiungere il loro pubblico di riferimento in modo più diretto ed efficiente, attraverso l'utilizzo di strumenti di targeting avanzati che consentono di raggiungere specifici segmenti di utenti e di offrire metriche di analisi che aiutino a valutare l'efficacia delle strategie di marketing intraprese;
- **accesso immediato alle notizie:** i *social media* hanno rivoluzionato la fruizione delle notizie, consentendo alle persone di ottenere informazioni in tempo reale. Tali piattaforme di *social media*, infatti, offrono la possibilità di seguire direttamente fonti di notizie, giornalisti ed organizzazioni, rimanendo costantemente aggiornati su quanto accade in tutto il mondo.

Ad oggi le tipologie di *social media* riconosciute sono sei e rientrano nella classificazione effettuata dagli stessi Kaplan e Haenlein (2010) sulla base di quattro variabili raggruppate a coppie: da un lato *social presence* e *media richness*, dall'altro *self-presentation* e *self-disclosure* (figura 6). In particolare, *social presence* riguarda la tipologia di contatto multimediale (fisico, visivo o acustico) che gli utenti possono avere, mentre *media richness* è relativo alla quantità di informazioni che possono essere trasmesse in uno specifico lasso di tempo. Tali variabili definiscono, quindi, l'immediatezza del mezzo e la sua abilità nel ridurre l'incertezza e l'ambiguità derivanti dalle informazioni presenti in esso. *Self-presentation* e *self-disclosure*, invece, sono riferiti ai dati ed alle informazioni che un utente è in grado di fornire e trasmettere su sé stesso, per cui riflettono l'abilità del mezzo nel facilitare la creazione di rapporti e relazioni tra gli utenti.

Figura 6 - Tipologie di social media secondo Kaplan e Haenlein (2010)

		social presence / media richness		
		Low	**Medium**	**High**
Self-presentation Self-disclosure	**High**	**Blog e Microblog** (Twitter)	**Social Networking Sites** (Facebook, Instagram)	**Virtual Social Worlds** (Second Life)
	Low	**Collaborative projects** (Wikipedia)	**Content Communities** (Youtube, Pinterest)	**Virtual Game Worlds** (World of Warcraft)

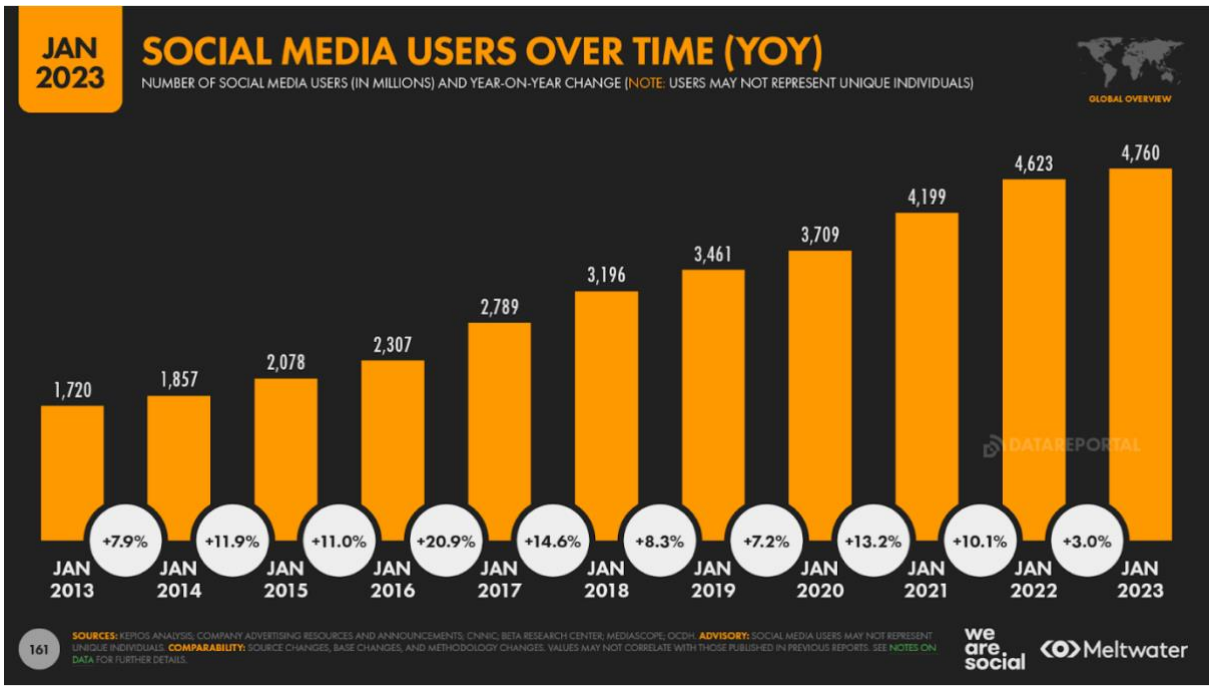
In particolare le 6 tipologie presenti in figura sono:

- ***blog* e *microblog*:** i blog sono piattaforme online che consentono agli utenti di condividere contenuti quali pensieri, opinioni e recensioni su diversi argomenti (es. WordPress). La possibilità di lasciare commenti ai post favorisce un discreto grado di interazione tra utenti. I microblog, invece, sono sempre piattaforme che consentono agli utenti di condividere brevi messaggi di testo in tempo reale, dando vita a comunicazioni istantanee (es. Twitter);
- ***social networking sites*:** comunemente detti *social network*, sono piattaforme che permettono agli utenti di creare un profilo personale, connettersi con amici, familiari e colleghi, e condividere contenuti quali messaggi foto e video (e. Facebook, Instagram, LinkedIn). Tali media facilitano la condivisione da parte degli utenti di elevate quantità di informazioni personali e la creazione di relazioni, talvolta strette, tra di essi;
- ***virtual social worlds*:** sono piattaforme che simulano mondi tridimensionali virtuali all'interno dei quali gli utenti assumono la forma di un *avatar* ed interagiscono tra loro suo tramite, come fossero in una realtà parallela (es. Second Life). In tal caso il livello di informazioni personali trasmesse e la *social presence* sono molto elevate;
- ***collaborative projects*:** sono siti web che favoriscono la creazione simultanea di contenuti da parte di diversi utenti, rappresentando pienamente il concetto di *User Generated Content* (UCG). Il contatto tra gli utenti di tali media è molto basso e allo stesso modo anche la quantità di informazioni personali trasmesse (es. Wikipedia);
- ***content communities*:** sono piattaforme attraverso le quali gli utenti hanno la possibilità di condividere contenuti di diverso tipo: foto, video, testo e presentazioni, limitando la trasmissione di informazioni personali a quelle legate agli stessi contenuti pubblicati (es. Youtube, Flickr, Pinterest);

- ***virtual game worlds*:** anch'essi come i *virtual social worlds* sono piattaforme all'interno delle quali gli utenti interagiscono sottoforma di *avatar*, ma in tal caso, riguardano giochi di ruolo online *multiplayer*. Il livello di interazione tra utenti è molto elevato a differenza della quantità di informazioni personali rilasciate da ciascuno di essi.

Al fine di comprendere pienamente la portata attuale del fenomeno dei *social media* e come esso si stia evolvendo, è possibile attingere ai dati forniti dall'agenzia We Are Social nel loro report globale annuale sul comportamento degli utenti online. Nel 2023, infatti, il numero totale di utenti attivi sui *social media* a livello globale ha raggiunto la cifra di 4.76 miliardi di utenti, che rappresenta il 59,9% della popolazione mondiale, con un incremento del 3% di utenti rispetto al 2022.

Figura 7 - Numero totale di utenti attivi sui social media nell'ultimo decennio

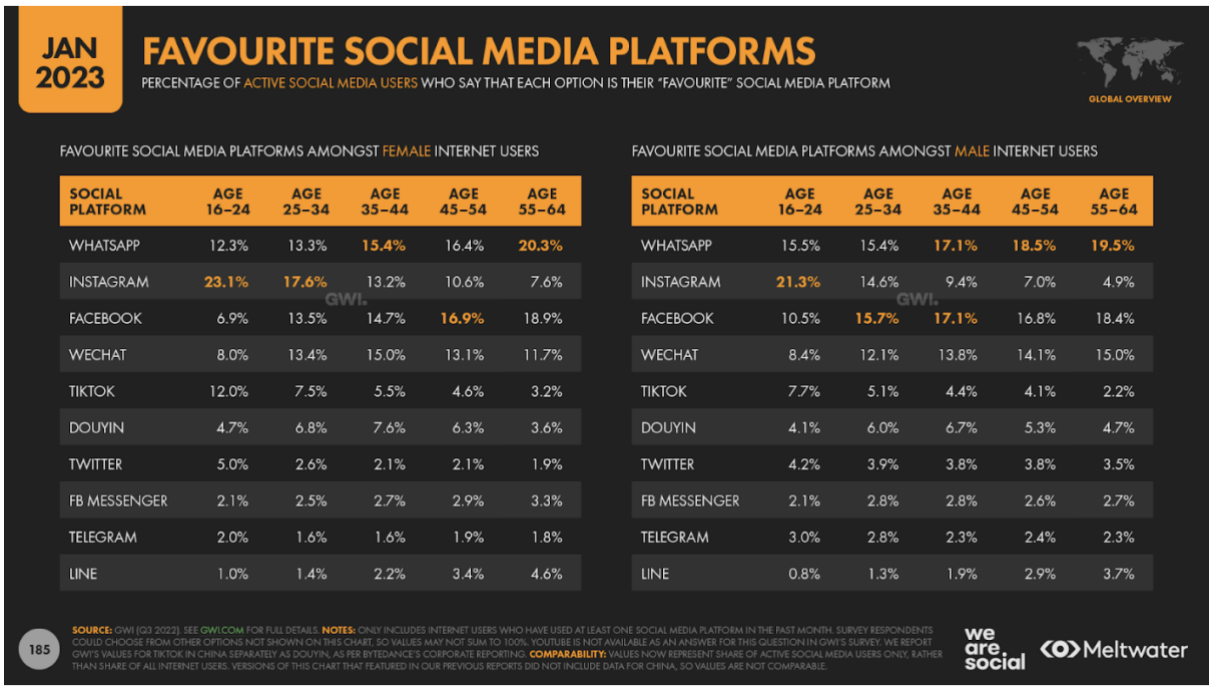


Fonte: We Are Social

Tra i *social media* più popolari, Facebook detiene il ruolo di leader, con oltre 2,9 miliardi di utenti attivi mensilmente (equivale quasi al 37% della popolazione mondiale), a seguire, al secondo posto, YouTube con più di 2,5 miliardi di utenti ed al terzo posto, a pari merito, Instagram e WhatsApp con 2 miliardi di utenti. Twitter, LinkedIn e Snapchat rientrano anch'esse tra le piattaforme più utilizzate, con centinaia di milioni di utenti ciascuno. A livello demografico, per gli utenti tra i 16 ed i 24 anni, Instagram rimane il *social med*ia più utilizzato, sia a livello femminile (23,1%), sia a livello maschile (21,3%). All'interno della stessa fascia di età le ragazze preferiscono TikTok (12,0%) a Facebook (6,9%), al contrario di quanto

avviene per i maschi i quali tendono ad utilizzare maggiormente Facebook (10,5%) rispetto a TikTok (7,7%). Un aspetto interessante è dato dalla preferenza, quasi totale, dai 35 anni in su per WhatsApp rispetto a tutti gli altri *social media*, a dimostrazione di come tutti coloro che non appartengono alla generazione dei "nativi digitali" preferiscono l'utilizzo di piattaforme più intuitive e facili da usare.

Figura 8 - Social media preferiti a seconda della fascia di età



Fonte: We Are Social

In sintesi, quindi, è possibile affermare che i *social media* abbiano assunto un peso rilevante ed una grande popolarità grazie alla loro capacità di connettere le persone, di facilitare la condivisione di contenuti, di promuovere l'interazione e di offrire opportunità di business da cui trarre vantaggio. La loro influenza sulla società e sulla cultura contemporanea è diventata sempre più significativa nel corso del tempo, tale da spingere molte persone ad utilizzare i *social media* come fonte di dati per una moltitudine di scopi differenti, grazie alla possibilità di accedere ad informazioni personali, notizie e contenuti generati dagli utenti stessi.

2.2 – *SOCIAL MEDIA DATA*: COSA SONO E PERCHÉ UTILIZZARLI

I dati generati sui *social media*, noti come *social media data*, si riferiscono all'insieme di informazioni e contenuti che vengono prodotti dagli utenti all'interno di piattaforme online (Fox, 2022). Essi rappresentano una preziosa risorsa per le aziende, i ricercatori e gli operatori di marketing poiché forniscono numerose informazioni relativamente all'interazione degli utenti, alle tendenze di consumo e alle opinioni degli individui. In particolare tali dati possono essere racchiusi all'interno di quattro categorie differenti (Juicer, 2021):

- **dati demografici e personali:** sono relativi ad età, genere, residenza, data di nascita, email, numero di telefono e molto altro. Tale tipologia di dati è quella più importante dal punto di vista della *privacy* e della tutela di un individuo: si è sempre restii a diffondere questa tipologia di informazioni proprio per questioni di riservatezza prediligendo, al contrario, dati che potenzialmente non mettono a rischio la propria identità (es. età e compleanno)
- **dati comportamentali:** riguardano comportamenti ed azioni frequentemente intrapresi da parte degli utenti come per esempio le attività di navigazione, i *click*, il tempo trascorso su una pagina, gli acquisti effettuati e il valore di tali acquisti e l'appartenenza ad un programma fedeltà. Tali dati forniscono una panoramica dettagliata del comportamento degli utenti e consentano di valutare l'efficacia delle strategie di marketing intraprese;
- **dati di *engagement*:** sono relativi alle interazioni sociali tra gli utenti all'interno di piattaforme *social*. Essi includono *like*, commenti, condivisioni, reazioni ai contenuti pubblicati dagli altri, visualizzazioni di pagina e siti web, tasso di *click-through* e molto altro. L'utilizzo di tali dati fornisce alle aziende informazioni in merito al coinvolgimento degli utenti e all'impatto che determinati contenuti, eventi e prodotti hanno sulla comunità;
- **dati attitudinali:** riguardano le emozioni e sensazioni degli utenti e misurano la percezione che essi hanno relativamente a determinati contenuti ed informazioni. A causa della loro natura prevalentemente soggettiva, tali dati vengono raccolti attraverso sondaggi, interviste, *feedback* da parte degli utenti, recensioni, *tweet* e messaggi testuali ed analizzati, successivamente, utilizzando il metodo della *sentiment analysis*.

Un aspetto molto importante relativamente ai dati provenienti dai *social media* è la loro natura. Essi, infatti, possono essere sia di carattere pubblico, sia privato, a seconda dell'accessibilità da parte di altri utenti della piattaforma (Fox, 2022). I dati pubblici comprendono tutte quelle informazioni che gli utenti condividono volontariamente con una vasta audience e possono, pertanto, essere visualizzate, lette ed utilizzate da chiunque abbia accesso alla piattaforma, ad esempio post e commenti. D'altra parte, dati considerati privati sui *social media*, comprendono

informazioni e contenuti che gli utenti scelgono di condividere solo con un gruppo ristretto di persone o con persone specifiche, per cui sono accessibili solo a coloro che hanno ricevuto l'autorizzazione da parte dell'utente per visualizzarli, ad esempio messaggi diretti e foto condivise con impostazioni di *privacy* limitate.

Tale differenza tra dati pubblici e privati ha implicazioni significative in termini di business e di analisi dei dati provenienti da *social media*. A livello competitivo, infatti, essendo generalmente accessibili con maggiore facilità, i dati pubblici forniscono una panoramica più ampia e rappresentativa, non solo del comportamento degli utenti, ma anche dei propri competitor e delle loro strategie facilitando, inoltre, eventuali confronti. L'utilizzo di dati di carattere privato richiede, invece, un approccio più attento e rispettoso della *privacy* degli utenti per cui talvolta può comportare processi più lunghi e complessi. In ogni caso, tale approccio non è esente dall'offrire vantaggi, in quanto facilita una comprensione più approfondita delle preferenze e dei gusti dei propri consumatori e utenti, così da sviluppare strategie di marketing mirate e offrire loro prodotti e servizi adatti.

L'utilizzo di *social media* come fonte di dati è aumentato significativamente poiché porta con sé una serie di vantaggi, soprattutto dal punto di vista economico e strategico. Essi, infatti, facilitano la raccolta dei dati rendendola più rapida, meno dispendiosa e meno invasiva rispetto all'utilizzo di metodi tradizionali quali questionari ed interviste. Inoltre, consentono di raccogliere e di immagazzinare enormi quantità di dati senza il problema di un archivio fisico in cui conservarli, come accade utilizzando sempre i metodi tradizionali. In aggiunta, i *social media* forniscono dati in tempo reale, costantemente aggiornati e offrono ampia visibilità a tutto ciò che è oggetto di discussione e dibattito su di essi, in quanto rappresentano gli strumenti di comunicazione maggiormente utilizzati al giorno d'oggi. Essi facilitano anche la creazione di *community* di utenti appassionati ad uno stesso prodotto, brand o argomento, all'interno delle quali si instaurano dibattiti preziosi per l'ottenimento di informazioni relative a pensieri, opinioni e percezioni.

A tale proposito, effettuare un'analisi relativamente alla percezione degli individui rispetto alla tutela della propria *privacy* e dei propri dati personali, utilizzando *social media data*, può rappresentare la soluzione adeguata. In *primis* perché, come detto in precedenza, i *social media* rappresentano una fonte importante di dati in tempo reale e quindi forniscono una panoramica ampia ed aggiornata del sentimento degli utenti rispetto al tema in questione.

In secondo luogo, i contenuti di carattere pubblico garantiscono, al contempo, sia un accesso diretto alle opinioni ed ai commenti degli utenti, sia l'identificazione delle tendenze di pensiero,

funzionali alla comprensione delle loro preferenze e delle loro reazioni agli stimoli generati dalle informazioni che ricevono dalla rete.

2.3 – LIMITI DI TALE APPROCCIO

L'utilizzo di dati provenienti da *social media* per effettuare studi ed analisi può, tuttavia, incontrare diversi limiti e *bias* che possono influenzare l'accuratezza e la rappresentatività dei risultati. In primo luogo, tali dati potrebbero fornire una visione limitata di quelle che sono le opinioni, i pensieri e gli stati d'animo di un utente, poiché dipendono dalla quantità e dalla tipologia di informazioni che esso decide di condividere (Investopedia, 2020). Alcuni utenti infatti possono decidere di non esprimersi direttamente, per esempio attraverso post o messaggi, ma limitandosi ad utilizzare strumenti quali il '*like'*, il '*retweet'*, o la ricondivisione, rendendo l'acquisizione di informazioni per effettuare analisi meno immediata o incompleta nel caso in cui mancassero effettivamente elementi ed informazioni importanti per la comprensione di un determinato argomento. Inoltre, i dati provenienti dai *social media* rappresentano solamente gli utenti attivi su tali piattaforme, tralasciando tutti coloro che non utilizzano quest'ultime o che non partecipano attivamente. Ciò comporta, quindi, inevitabilmente una visione parziale o distorta della realtà (***bias* di selezione**). Un ulteriore aspetto che potrebbe causare distorsioni è dato dagli stessi *social media*. Sebbene, infatti, essi rappresentino strumenti comunicativi ampiamente diffusi, ciascuno possiede utenti con caratteristiche demografiche differenti. Alcune piattaforme, sono più popolari tra i giovani come per esempio Instagram e TikTok mentre altre tra gli adulti, come per esempio Facebook e WhatsApp (***bias* demografico**). Sempre a livello demografico, è possibile ottenere distorsioni nei risultati di un'analisi in quanto ci sono soggetti che utilizzano le piattaforme *social* per scopi differenti, per esempio c'è chi utilizza Twitter solo come fonte di notizie e dati in tempo reale mentre c'è chi lo utilizza per connettersi con altri utenti, esprimersi e creare contenuti (Olteanu et al. 2019).

Anche l'aspetto geografico gioca un ruolo importante nel determinare la presenza o meno di distorsioni. Al riguardo, i dati provenienti dai *social media* potrebbero essere influenzati dalla provenienza geografica degli utenti, poiché, in alcune aree, determinati argomenti risultano maggiormente sentiti e ricondivisi rispetto ad altre in cui, invece, hanno un ruolo marginale o sono addirittura tabù. Ciò potrebbe comportare una distorsione dei risultati a favore di posizioni ed opinioni che riflettono, appunto, le credenze e le tradizioni di una determinata zona (***bias* geografico**). Infine, dati provenienti dai *social media* potrebbero essere influenzati da *trend,* eventi e circostanze proprie di un determinato momento o arco temporale (***bias* temporale**). A livello temporale, però, per evitare possibili *bias* è importante tenere in considerazione anche

la natura del fenomeno in analisi, ossia se esso rappresenti un fenomeno stagionale, periodico o improvviso. Tali differenze, infatti, possono comportare un utilizzo differente dei *social media* da parte degli utenti, stimolando così una creazione di dati informazioni e contenuti differente dal solito. La presenza di tanti dati che appaiono più volte all'interno di un *dataset* sottoforma di duplicati può portare alla creazione di un effetto di **ridondanza**, per esempio attraverso *retweet* e contenuti ricondivisi. Tale effetto può contribuire all'ottenimento di risultati talvolta distorti anche se, quando accade, esso rappresenta un segnale ossia, che il pensiero o l'argomento in questione riveste un ruolo centrale all'interno della discussione e quindi ha importanza (Olteanu et al. 2019). Infine, un ultimo limite che è possibile incontrare analizzando dati provenienti da *social media* è relativo alla natura stessa dei dati. Esistono, infatti, dati che rappresentano delle metriche fattuali, ossia più indicative ed utili, in quanto basati su informazioni sicure e concrete, ad esempio *like*, commenti e condivisioni (Fox, 2020); d'altro canto, invece, esistono dati che rappresentano solamente metriche di stima, poiché sono comunque indicative, ma meno accurate ed affidabili rispetto a quelle fattuali poiché basate su approssimazioni, ad esempio le *impression* e la *reach* (Fox, 2020). I dati di stima, solitamente, si utilizzano quando quelli fattuali non sono disponibili o risultano incompleti. Pertanto, nell'effettuare analisi con tali tipologie di dati è importante tenere in considerazione che non tutte le metriche potrebbero essere egualmente indicative ed utili ai fini del raggiungimento di un risultato.

Tenuto conto di quanto precede e considerato che l'analisi presente all'interno di tale elaborato verrà eseguita utilizzando dati provenienti da *social media*, in particolare *tweet* contenenti *#privacy*, ci si aspetta di ottenere una visione parziale della realtà, poiché tali *tweet* riflettono solamente le opinioni e le percezioni degli utenti attivi su Twitter. Inoltre, il fatto che Twitter sia un social network molto più diffuso nella fascia di popolazione giovane, lascia intendere che la percezione relativa alla tutela della propria *privacy* e dei propri dati possa essere distorta verso il positivo in quanto, essendo i giovani nati come "nativi digitali", essi percepiscono l'utilizzo della tecnologia e dei numerosi dispositivi digitali, come un qualcosa di normale, senza i timori e le preoccupazioni tipiche degli adulti. A livello di possibili distorsioni temporali, ci si aspetta che i *tweet* relativi al tema della *privacy* possano essere influenzati da qualche avvenimento o circostanza attuale, essendo comunque tale tema profondamente legato con ogni ambito della vita quotidiana e, quindi, ampiamente discusso e dibattuto su base giornaliera. Per tale motivo, la presenza di un eventuale *bias* temporale non inciderebbe sul risultato dell'analisi, anzi, contribuirebbe a fornire un'ulteriore nuova sfaccettatura legata all'argomento della *privacy* e della protezione dei dati personali.

Utilizzando Twitter come fonte di dati, ci si aspetta, inoltre, di ottenere un possibile effetto di ridondanza dato dalla presenza di numerosi *retweet* all'interno del *dataset*. Tale effetto potrebbe contribuire a distorcere i risultati sia verso una percezione positiva, sia negativa; in ogni caso, a livello generale, contribuirebbe a sottolineare l'importanza e la centralità, all'interno della discussione, dell'argomento preso in analisi, sottolineando la presenza di numerosi utenti che condividono lo stesso sentimento.

A fronte delle suddette limitazioni e possibili distorsioni, la scelta di utilizzare ugualmente Twitter come fonte di dati per l'analisi ricade sull'elevata percentuale di profili pubblici che facilitano l'accessibilità ai dati, rendendo tale piattaforma *"una vera e propria miniera di dati liberi"* (Ceron et al, 2014). Inoltre, la possibilità di indicizzare il proprio *tweet* attraverso l'uso di un *hashtag* consente di inserirlo all'interno di una discussione generale in cui i partecipanti si uniscono per il solo fatto di condividere l'argomento e non necessariamente per la conoscenza personale diretta, data anche dalla provenienza geografica. Ciò contribuisce a rendere le discussioni presenti su Twitter più ampie e di carattere generale, ma ugualmente ricche di sfumature, in quanto includono diversi punti di vista.

2.4 – IL PROBLEMA DELLA PRIVACY

Il legame che intercorre tra *social media*, analisi dei *social media data* e *privacy* risulta oggi, più che mai, attuale, complesso e rilevante.

Attuale, poiché viviamo in una società in cui l'utilizzo dei *social media*, di Internet e dei dispositivi digitali ha assunto grande importanza e pervasività, rendendo quasi impensabile immaginare di condurre una vita quotidiana senza il loro utilizzo. Inoltre, in seguito ai numerosi scandali relativi a violazioni della *privacy* e dei dati personali, avvenuti nel corso degli ultimi anni, è aumentata l'attenzione e la consapevolezza degli individui nei confronti di tali fenomeni.

Complesso, poiché la tutela precisa ed accurata sia della *privacy*, sia dei dati personali degli utenti è resa quotidianamente più ardua dalla stessa mole di dati generati sui sistemi dagli utenti. I miliardi di utenti iscritti hanno, infatti, la possibilità di collegarsi e raffrontarsi con altri, pubblicando contenuti, esprimendo opinioni e pensieri, prendendo parte a dibattiti e discussioni, creando svariate occasioni di socializzazione e confronto a livello virtuale. La complessità del legame in questione risiede, peraltro, anche nella condivisione dei *social media data* con soggetti di terze parti (Castro, 2010), quali, ad esempio, le agenzie pubblicitarie. In particolare, ciò può suscitare preoccupazioni e timori da parte degli utenti in merito alla trasparenza e al controllo dei propri dati. In termini di trasparenza, infatti, è basilare che gli utenti sappiano quali dei loro riferimenti personali vengano condivisi, con chi, come e perché al fine di consentirgli di prendere decisioni consapevoli sul loro agire, ed infondendogli un senso di sicurezza e fiducia.

Rilevante, per il fondato timore che i dati raccolti all'interno anche dei *social media* possano essere utilizzati per motivi impropri, quali furti di identità, truffe, *phishing* e altre forme di criminalità informatica che potrebbero avere ripercussioni significative nella vita personale e professionale degli utenti, influenzando anche la loro reputazione. Alla luce di un futuro che si prospetta sempre più digitale, tecnologico ed automatizzato la protezione della *privacy* e dei dati personali si rivela, quindi, essenziale per preservare il concetto di autonomia e controllo degli individui sulla propria vita digitale. Altro aspetto connesso con la rilevanza, che merita un cenno, per completezza di trattazione, è l'ambito etico. A sottolinearne la particolare valenza, laddove questa non fosse già implicita nell'etica professionale del singolo operatore/funzionario/dirigente, si registra la recente pubblicazione di una guida denominata *"Data Ethics – The Rise of Morality in Technology"* (World Federation of Advertisers, 2020) finalizzata a fornire alle aziende indicazioni e consigli relativi a come attuare un approccio etico. Secondo la suddetta guida sarebbero quattro i principi chiave che sottendono ad un codice di condotta etico: Rispetto, Equità, Responsabilità e Trasparenza. È possibile leggerli da sinistra

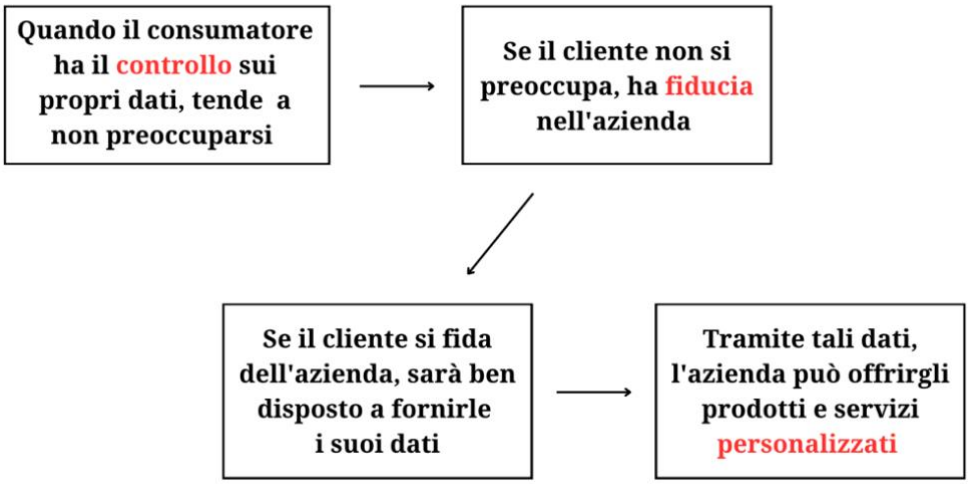
a destra o viceversa, ritornando automaticamente a concetti già enunciati nelle pagine precedenti, ad ulteriore dimostrazione di come il connubio tra *privacy* ed etica rappresenti la chiave per la società digitale del futuro, ossia una società in cui tecnologia e progresso continuino a facilitare e migliorare la vita dell'uomo nel rispetto dei suoi diritti.

In merito, è opportuno non sottovalutare che tutto quello che viene pubblicato sui *social media* non può essere considerato di carattere privato in quanto l'obiettivo della condivisione è proprio quello di dare visibilità a tali contenuti agli occhi degli altri (Flint, 2009). Contenuti di carattere pubblico sono, quindi, visibili ed accessibili a chiunque e per questo potenzialmente soggetti a rischi. Quanto sopra citato non deve rappresentare per gli utenti un limite alla propria espressione, ma solo un monito a condividere veramente quello che essi desiderano e che ritengono non dannoso per la propria identità. Una delle motivazioni principali dell'insorgenza di problemi legati alla *privacy* deriva dal fatto che in fase di registrazione ad una piattaforma, laddove non si sia obbligati ad "accettare" per poter procedere, per questioni di pigrizia, di facilità e di rapidità, gli individui prestano poca attenzione a quello che accettano, non acquistando consapevolezza in merito a quanto potrebbe accadere ai loro dati personali. Ciò dimostra l'esistenza di una vera e propria dicotomia tra le intenzioni e le azioni dei clienti nei confronti della *privacy* (Kokolakis, 2017). Si tratta del cosiddetto 'Paradosso della *Privacy*' (o *privacy paradox*) che pone in relazione la paura, le preoccupazioni e i freni degli individui nel condividere informazioni con il loro effettivo comportamento. Molti, infatti, sono restii al rilascio di dati ed informazioni per il timore di un loro utilizzo improprio, ma nonostante ciò si dimostrano comunque disponibili alla loro diffusione, se ricompensati. Tale dicotomia diventa doppiamente paradossale nel momento in cui gli individui, contrari alla condivisione di dati a fini commerciali, si rivelano essere utenti molto attivi sui *social network*, dove qualsiasi azione comporta la creazione di tracce digitali. Nella letteratura è possibile individuare una serie di elementi che potrebbero potenzialmente spiegare tale paradosso. In *primis*, la poca consapevolezza e conoscenza delle possibili conseguenze derivanti dalla condivisione di informazioni (Pew Research Center, 2019), unitamente alle non corrette valutazioni rischi/benefici della diffusione di informazioni da parte dei consumatori. Essi, infatti, di fronte agli immediati vantaggi derivanti dalla tecnologia perdono di vista i potenziali rischi per i propri dati (Acquisti, 2004). Infine, l'illusione di avere un controllo sui propri dati derivanti dalla possibilità di scegliere quali condividere e quali no (Brandimarte, Acquisti, Loewenstein, 2013).

Considerato quanto precede e tenuto conto della valenza dei *social*, è possibile effettuare analisi sui dati provenienti da essi per estrarre preziose informazioni in merito al comportamento degli

utenti, alle loro opinioni e alle tendenze di mercato, da utilizzare successivamente per prendere decisioni aziendali, formulare strategie di marketing ed instaurare rapporti di fiducia con gli individui. Ciò giustifica il forte interesse delle imprese verso la raccolta e l'analisi dei dati dei consumatori. Anche in tal caso, l'idea che i propri dati vengano analizzati può generare negli utenti sentimenti contrastanti derivanti dai possibili rischi e benefici a cui essi potrebbero essere sottoposti. Al riguardo, al fine di ovviare a tale problema, le imprese dovrebbero trasmettere ai propri clienti un'immagine di professionalità ed efficienza che contribuisca a rassicurarli, donandogli un senso di protezione. Inoltre, la maggior parte delle aziende che operano con i *social media data* hanno a disposizione tre leve attraverso cui agire migliorando la percezione della *privacy* nei consumatori e facilitare, così, il rilascio di dati ed informazioni utili in modo più consapevole e sicuro: fiducia, personalizzazione e controllo (Martin, Murphy, 2017). La **fiducia** rappresenta un elemento fondamentale per la creazione di rapporti solidi e duraturi tra le parti. La sua presenza all'interno di tali relazioni, è cruciale perché contribuisce ad attenuare l'impatto delle paure e delle preoccupazioni degli individui. Inoltre, la fiducia dei consumatori nei confronti di un'azienda è strettamente collegata all'immagine che essi hanno di quest'ultima, per cui riuscire a guadagnare o aumentare la loro fiducia ha un forte impatto anche sulla reputazione aziendale. Quando c'è fiducia da parte dei consumatori si innesca in loro il desiderio di condividere tale esperienza positiva con parenti, conoscenti e amici, dando vita ad un passaparola importante per l'azienda in termini di vendite e di reputazione. La **personalizzazione** riguarda la capacità delle imprese di fornire ai consumatori prodotti e servizi che rispecchino i loro gusti e preferenze. Al riguardo, le aziende necessitano di raccogliere grandi quantità di dati ed informazioni di quest'ultimi, ma ciò può avvenire nel momento in cui essi sono disposti a condividere le loro informazioni per ricevere un trattamento personalizzato. La fiducia, quindi, riveste un ruolo centrale anche ai fini della personalizzazione, poiché contribuisce ad instillare nei consumatori quel senso di protezione e sicurezza che favorisce una maggiore disponibilità al rilascio di dati. Fiducia e personalizzazione sono due elementi strettamente collegati fra loro. Infine, La terza leva utilizzata dalle imprese per intervenire sulla percezione della *privacy* è il **controllo**, ovvero la possibilità data ai consumatori di decidere quali dati condividere. Come accennato precedentemente, l'illusione di avere il controllo sulla diffusione delle proprie informazioni suscita in essi una preoccupazione minore per la loro *privacy*. Al contrario, invece, in assenza di controllo i consumatori si sentono vulnerabili, e per questo meno propensi a condividere i loro dati con chi manca della loro fiducia.

Figura 9 - Schema riassuntivo del legame tra fiducia, personalizzazione e controllo



In sintesi, la *privacy* è un tema cruciale, che richiede ai vari livelli organizzativi, impegno ed iniziative efficaci da parte delle istituzioni, dei gestori dei *social media*, delle aziende e, soprattutto, degli utenti stessi. In merito, infatti, sebbene nel corso degli anni i *social media* e le istituzioni abbiano adottato politiche di sviluppo, *software* di gestione e legislazioni specifiche per tutelare la sicurezza degli individui e dei loro dati personali in rete, esistono ancora molti utenti che non leggono e non prestano attenzione a queste politiche e ai di termini di utilizzo. La soluzione di tale divario, originato dall'intento di proteggere gli utenti laddove essi stessi mancano della necessaria consapevolezza ad autotutelarsi, sarà una delle sfide più importanti per giungere ad una società digitale, sostenibile ed eticamente rilevante.

3. ANALISI EMPIRICA SUI DATI DI TWITTER

"I dati sono diventati il quarto fattore produttivo, dopo i classici terra, lavoro e capitale." (Vincenzo Cosenza)

3.1 – RACCOLTA E DESCRIZIONE DEI DATI

I dati utilizzati per effettuare le analisi presenti all'interno di questo elaborato provengono da Twitter, una piattaforma online che consente di comunicare direttamente pensieri, emozioni, sensazioni, notizie e molto altro ancora attraverso brevi messaggi chiamati 'tweet'. I *tweet*, infatti, sono comunicazioni di dominio pubblico, composte da un numero limitato di caratteri, accessibili a tutti gli utenti collegati e per questo scelti come fonte di dati per l'analisi. Inoltre, all'interno di Twitter è possibile indicare l'argomento principale di un *tweet* attraverso l'utilizzo di parole chiave precedute dal cancelletto (#) detto *hashtag* (es. #terremoto).

Sulla base di quanto precede, il reperimento dei dati in questione è stato effettuato utilizzando l'*Application Programming Interface* (API) ufficiale di Twitter, tramite il pacchetto 'rtweet' del software 'R'. Sono stati presi in considerazione tutti i *tweet* in lingua inglese contenenti #*privacy*, compresi nell'intervallo di tempo compresi tra il 12 e il 18 marzo. Al riguardo, tuttavia, il recente acquisto della piattaforma da parte dell'imprenditore Elon Musk ha portato con sé l'introduzione di numerosi cambiamenti riguardanti diversi aspetti, tra cui anche l'API. In merito, quale conseguenza immediata, è stata momentaneamente revocata agli utenti che utilizzano l'API la possibilità di inserire una serie di filtri, quali la geolocalizzazione e la selezione dell'arco temporale durante l'estrazione dei *tweet*, motivo per cui i dati del seguente elaborato risultano condizionati.

3.2 – ANALISI ESPLORATIVA

Il dataset preso in considerazione è composto da 13.075 *tweet*. La normalizzazione del testo di tali *tweet* è stata inizialmente effettuata in modo manuale, rimuovendo gli URL ed altri caratteri che non risultavano informativi. Successivamente, sono stati rimossi ulteriori elementi quali numeri, punteggiatura, congiunzioni, articoli, preposizioni e spazi bianchi.

L'analisi esplorativa è stata condotta attraverso l'impiego di tre diverse tecniche. In particolare:

- analisi delle parole più frequenti (*word cloud*);
- analisi degli *hashtag* più frequenti;

- analisi dei *tweet* con maggior numero di *retweet* e *like*.

Dal punto di vista semantico, l'analisi esplorativa prevede l'utilizzo della *word cloud*, uno degli strumenti più comunemente utilizzati per effettuare analisi su dati provenienti da *social network.* In particolare, la *word cloud* corrisponde alla rappresentazione grafica delle parole impiegate con maggior frequenza. L'importanza di ciascuna parola si evince dalle dimensioni e dai colori con cui vengono raffigurate. Inoltre, è possibile, al fine di facilitare l'analisi, raggruppare le parole più frequenti in macro categorie. In particolare, in questo caso, le parole maggiormente utilizzate sono state ricondotte a quattro macro categorie collegate al tema della *privacy*, che dimostrano quali siano gli ambiti in cui tale aspetto assume rilevanza. Tali categorie sono raffigurate nella tabella sottostante, posizionate in sequenza decrescente in base alle dimensioni delle parole nella *word cloud*.

Figura 10 - Word cloud delle parole più frequenti usate nei tweet contenenti #privacy

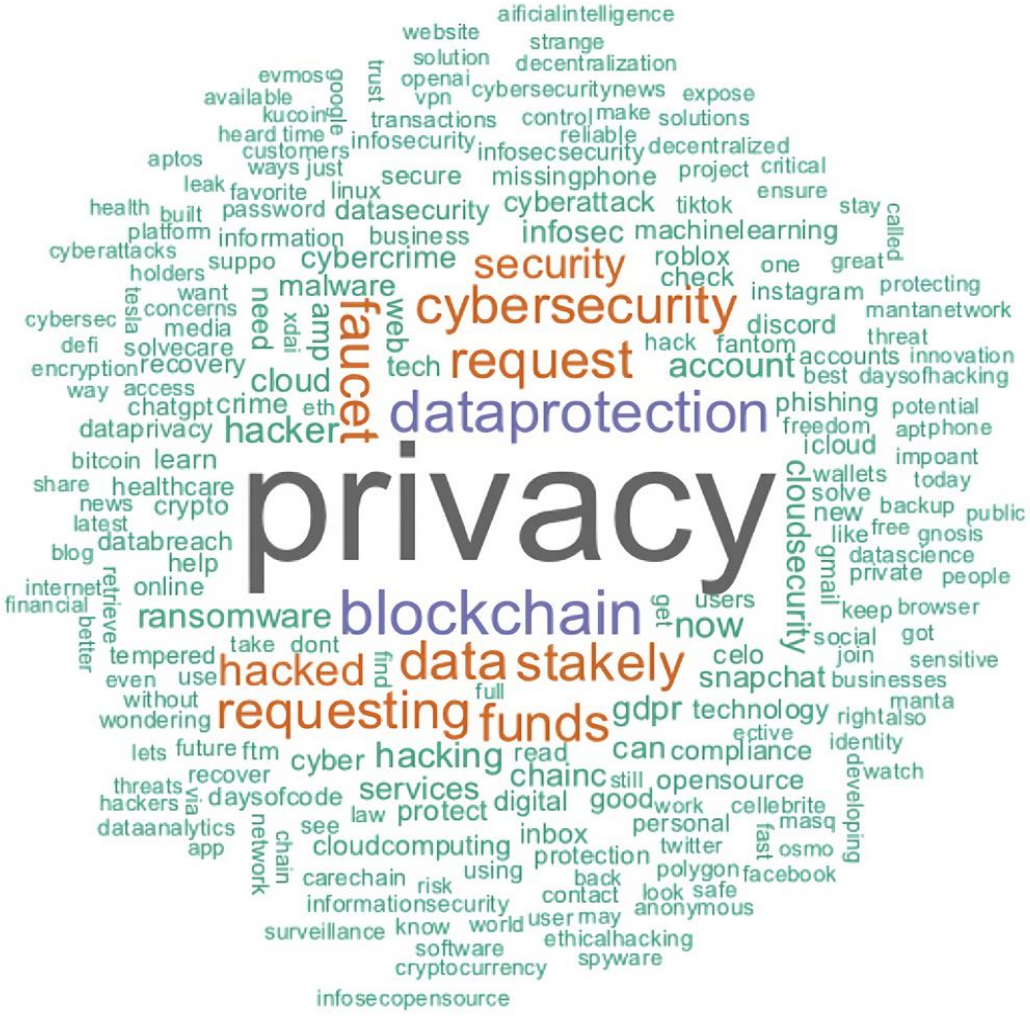


Tabella 2 - 4 macro categorie a cui sono riconducibili le parole presenti nella word cloud

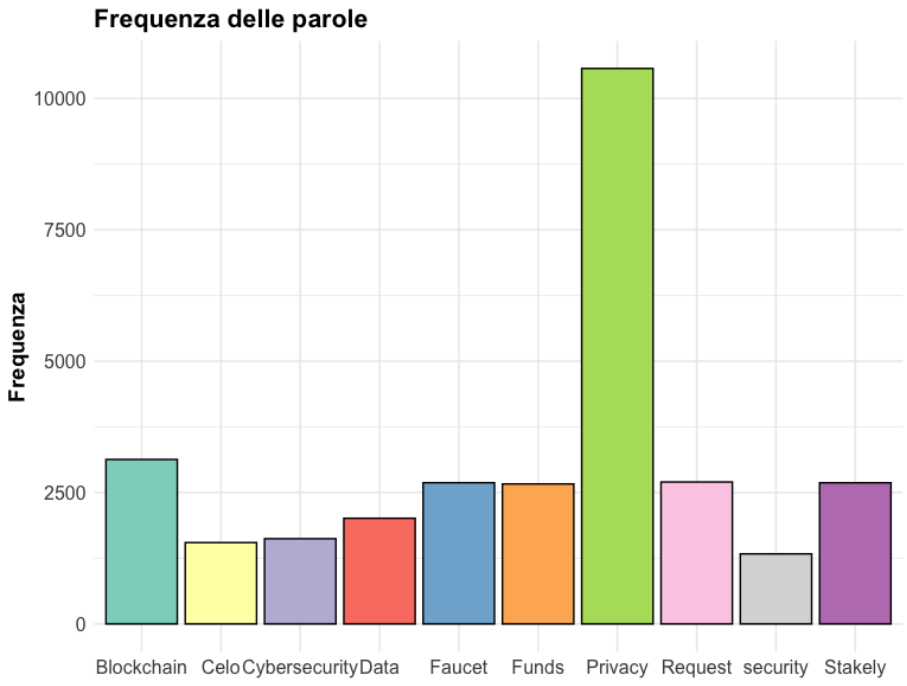
Categoria	Parole esemplificative della categoria
BLOCKCHAIN/CRIPTOVALUTE	blockchain, faucet, stakely, request, cryptocurrency, bitcoin
SICUREZZA/PROTEZIONE DATI	data protection, security, cybersecurity, ethical hacking, data privacy, data security, cloudsecurity, encryption, gdpr
POTENZIALI RISCHI	hacked, surveillance, cyberattack, malware, leak, data breach, phishing, hacker
SOFTWARE E STRUMENTI	data analytics, artificial intelligence, Facebook, Instagram, Tik Tok, Roblox, Chatgpt, machine learning

I *tweet* considerati contengono tutti *#privacy* per cui, oltre alla parola *privacy*, a seguire, è possibile trovare due parole in viola e numerose in arancione delle quali, tranne alcune, sono tutte relative al tema della *blockchain*. Ciò dimostra un forte coinvolgimento da parte degli utenti per una tecnologia che in futuro potrebbe essere sempre più diffusa ed adottata, in quanto garantisce una maggiore sicurezza in termini di *privacy* e di dati personali i quali, una volta archiviati e protetti dalla crittografia, sono accessibili solo agli utenti della *blockchain* stessa, non sono eliminabili, ma soprattutto non sono modificabili senza il consenso di tutti gli utenti.
La seconda e terza macro categoria presenti all'interno della tabella contengono termini legati ai temi della sicurezza/protezione dati ed ai rischi per gli utenti. La maggiore frequenza relativa alla prima categoria dimostra che, sebbene la consapevolezza rispetto ai possibili rischi sia elevata, l'attenzione degli utenti è focalizzata al desiderio ed alla necessità di tutelare sé stessi. In particolare, la diffusione capillare dei *social media* e della cosiddetta modalità online, unitamente alla consapevolezza dei rischi associati all'uso della rete, hanno instillato l'idea che in un'ottica costi/benefici, i benefici apportati dalla tecnologia comportino automaticamente una serie di potenziali pericoli, ritenuti accettabili a patto che sia fatto il possibile per tutelare i cittadini/consumatori. Alle aziende, quindi, conviene investire e puntare sulla sicurezza tanto quanto sul marketing e sulla produzione, poiché anche loro sono potenzialmente soggette agli stessi rischi degli individui. Proteggere sé stesse è il primo passo per proteggere i propri clienti ed instaurare, così, delle relazioni profittevoli fondate su fiducia e rispetto reciproco.
Infine, tra le parole più frequentemente utilizzate, presenti nella *word cloud* in verde ed inserite al quarto posto della tabella, ci sono proprio quelle riconducibili alla categoria dei *software*. Al

riguardo, è possibile affermare che software e strumenti rappresentino applicazioni della tecnologia entrate a far parte della quotidianità degli individui e divenute per le aziende elementi importanti e strategici. In particolare, attraverso quest'ultime, le aziende hanno la possibilità di offrire ai partecipanti un'esperienza unica e coinvolgente, mantenere viva la relazione con essi, coltivandola, e sfruttando a pieno il vero valore dei dati per creare prodotti e servizi che rispecchino e soddisfino i loro gusti. All'interno di tale categoria, quindi, è possibile trovare parole legate al mondo dei *social network* e delle piattaforme come Instagram, Facebook e Roblox[36], ma anche e soprattutto legate alla *data analytics* e all'intelligenza artificiale.

Per completezza di trattazione, di seguito è riportato un grafico riepilogativo raffigurante le 10 parole più frequentemente utilizzate e le loro specifiche frequenze. Tale grafico è da considerarsi complementare alla *word cloud* per una comprensione più approfondita delle principali tematiche e tendenze legate all'argomento della *privacy*. Tali parole, infatti, sono tutte coerentemente riconducibili alle quattro categorie identificate precedentemente, analizzando la *word cloud*.

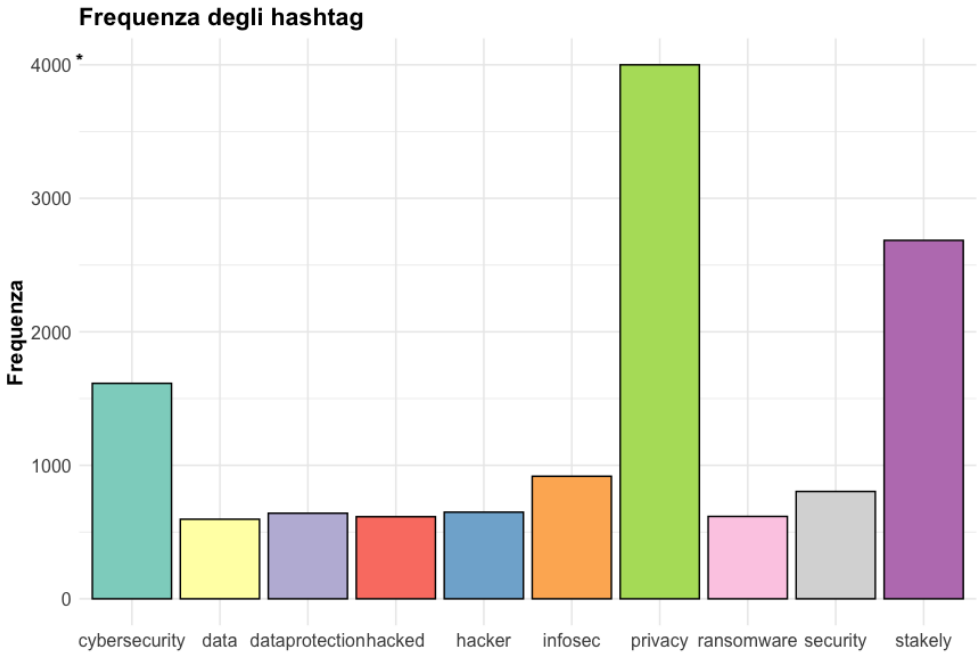
Figura 11 - Grafico raffigurante le 10 parole più frequentemente utilizzate nei tweet contenenti #privacy



[36] Roblox è una piattaforma di gioco online che funge anche da sistema per lo sviluppo e la creazione di videogiochi.

La seconda tecnica è l'analisi degli *hashtag* più utilizzati. Essi, infatti, consentono di etichettare e categorizzare i contenuti sui *social media*. Nello specifico, su Twitter, gli *hashtag* contribuiscono ad inserire il proprio *tweet* all'interno di un dibattito collettivo tra tutti gli utenti relativo ad uno specifico tema, per cui frequenze elevate di uno stesso *hashtag* stanno ad indicare che l'argomento è molto popolare e dibattuto. Il grafico sottostante riporta i 10 *hashtag* più frequentemente presenti nei *tweet* appartenenti al dataset considerato.

Figura 12 - Grafico raffigurante i 10 hashtag più frequentemente utilizzati nei tweet contenenti #privacy



*Il range della frequenza è stato ridotto per visualizzare meglio le minime differenze presenti, pertanto la frequenza di #privacy sarebbe superiore a 9000 unità

In particolare, da esso si evince che la maggior frequenza è attribuita ad *hashtag* che hanno per sfondo comune il campo della sicurezza e tutela della *privacy*. Si passa, infatti da #cybersecurity ad #infosec, da #dataprotection ad #privacy, fino ad #security. All'interno del grafico, inoltre, sebbene con frequenza molto minore, si trovano anche hashtag relativi al mondo delle criptovalute e della *blockchain*. Tale situazione, però, si differenzia da quella presente nel grafico delle parole più frequenti (Figura 11), dove ad essere in maggioranza sono proprio termini legati a quest'ultimo tema. Il contrasto tra i due appare curioso, ma può essere spiegato considerando che, in generale, gli individui sono sempre molto restii a rivelare le proprie intenzioni quando realizzano di avere concretamente qualcosa da perdere, come ad esempio il guadagno sugli investimenti, ovvero sull'incremento della propria capacità economica. È per

questo che preferiscono dibattere il tema associato della *privacy* e della tutela dei propri dati personali (in questo caso dei propri investimenti), per dedurre dalle informazioni che ricevono quelle più interessanti al loro caso particolare.

La terza tecnica è quella relativa ai *retweet*, ovvero ai *tweet* più riproposti, ed ai *like*. In particolare, un aspetto interessante dei dati in questione è la grande percentuale di *retweet* presenti nel dataset (47% del totale). Questo indica come, relativamente ad uno specifico argomento, molti utenti tendano a ricondividere il pensiero di altri profili piuttosto che esporsi dichiarando apertamente la propria opinione. Ciò non toglie, tuttavia, l'aspetto della responsabilità, poiché *'retweettando'* si esprime la volontà di diffondere il contenuto, appropriandosene intellettualmente e, appunto, divenendo responsabili per quanto scritto dall'altro utente.

Lasciare un 'mi piace' (o *like*) ad un post o ad un commento, significa esprimere il proprio gradimento in merito. Il 'mi piace' ha un significato analogo e può indicare apprezzamento per il contenuto del post o del commento, ma anche interesse verso di esso e coinvolgimento. Per cui, anche gesti che possono apparire molto semplici come un *retweet* o un *like*, sono in realtà portatori di informazioni preziose relativamente a gusti, preferenze, opinioni e pensieri degli individui. Nello specifico, all'interno delle tabelle sottostanti sono riportati i cinque *tweet* con il maggior numero di *retweet* e i cinque *tweet* con il maggior numero di *like*. In merito, si può facilmente notare come i *retweet* siano complessivamente molto più elevati dei *like* e quasi tutti riferiti al tema della *blockchain* e delle criptovalute dal punto di vista della sicurezza delle transazioni e dei dati. I *tweet* con più 'mi piace' riguardano, invece, per la maggior parte, l'argomento della *privacy*, ma ne sono presenti anche due dai temi importanti, uno dei quali, quello con il maggior numero di *like*, riferito all'ambito della sicurezza, mentre l'altro, quello con il minor numero di *like*, riferito alla questione etica. Tale aspetto risulta in linea con quanto emerso precedentemente dal confronto fra le parole e gli *hashtag* più frequenti.

Tabella 3 - Riepilogo dei 5 tweet con più retweet

TWEET PIÙ RETWEET	NUMERO RETWEET
Ruby Protocol ♦ @RubyProtocol ❤️ Crypto world gives us nothing but love. 🚀 With the tremendous growth in all our channels and, most importantly, in Ruby Connect. ⏰ It's time to launch a #Privacy Contributor #Giveaway Program to thank all that support our cause.	11.338
Tusima @TusimaNetwork 🔥We are excited to announce that Tusima Testnet is live now! When you complete the test, you will be rewarded a TusimaDAO SBT Here is the test guide link.medium.com/fDP0cZxnKtb 🎁MANY EVENTS and REWARDS for #Testnet participants! #zkRollup #Layer2 #privacy #Blockchain #trias	7853
Mysk @mysk_co We confirm that iOS 16 does communicate with Apple services outside an active VPN tunnel. Worse, it leaks DNS requests. #Apple services that escape the VPN connection include Health, Maps, Wallet. We used @ProtonVPN and #Wireshark. Details in the video: #CyberSecurity #Privacy	7197
OMNIA Protocol o.O @omnia_protocol ⚡#Metaverse, #Defi, #NFTs, #GameFi - over 200B+ markets are still relying on centralized infrastructure. ✅ OMNIA eliminates single points of failure. ✅ Revenue streams. ✅ #Privacy-first access to the #blockchain. $OMNIA launching soon 🚀 See more: omniatech.io	2644
Partisia Blockchain @partisiampc In our second edition of Partisia Blockchain Tech Talks, our leadership team provides some incredible insights into why we built Partisia Blockchain and how it solves the #blockchaintrilemma #Privacy #Scalability #Interoperability #MPC	585

Tabella 4 - Riepilogo dei 5 tweet con più like

TWEET PIÙ LIKE	NUMERO LIKE
Hedera ✓ @hedera ··· #Hedera is proud to partner with @IdemiaGroup on their mission to provide secure, customizable, & accessible solutions to Central Banks exploring #CBDC - unlocking the potential of connected #web3 economies while preserving user #privacy. Learn more idemia.com/secure-offline...	794
International Human Rights Foundation ✓ @Declaracion ··· 6. We recommend that citizens take self-protection measures: reduce the chances of their faces being recorded or photographed by violent agents, avoid being followed, go accompanied when leaving protests and avoid going directly home. #Protection #Security #Privacy	186
Tutanota @TutanotaTeam ··· The EU Commission is planning what Apple stopped after a huge backslash: Turning your own device into a surveillance machine via client-side scanning. Stop #ChatControl now! 💪 Join the fight for #privacy:tutanota.com/blog/posts/csa...	174
Proton ✓ @ProtonPrivacy ··· Internet shutdowns and #censorship are rampant worldwide. This #WorldDayAgainstCyberCensorship, we're kicking off our inaugural mini-doc series with a video interviewing #privacy activists about what online censorship is and how to circumvent it.	152
Proton ✓ @ProtonPrivacy ··· Services like #ChatGPT have the power to transform our society. They can also decimate our online #privacy. Ethics cannot be an afterthought for Microsoft when it comes to such a powerful new tool. Thanks to @carissaveliz for sharing this article.	141

3.3 – ANALISI DEL SENTIMENTO

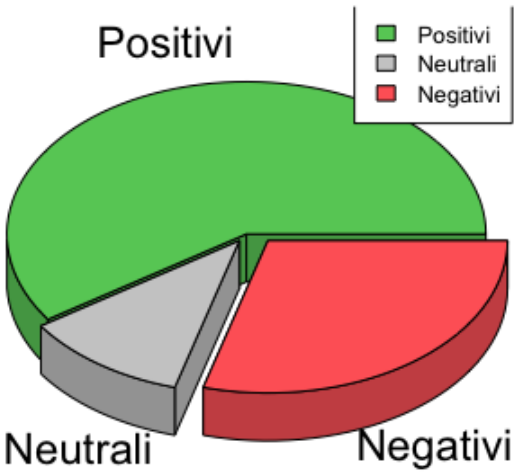
In questa seconda parte del capitolo, l'attenzione è focalizzata ad indagare e comprendere la percezione degli individui relativamente alla tutela della propria *privacy*. Al riguardo, è utile effettuare una *sentiment analysis* che prevede l'analisi dei *tweet* presenti all'interno del *dataset*, dal punto di vista dei sentimenti, dei giudizi e delle opinioni degli utenti.

L'obiettivo principale è quello di identificare se il testo trasmette un sentimento positivo, negativo o neutro.

Uno dei metodi più utilizzati per effettuare la *sentiment analysis* prevede di considerare l'intero testo di un *tweet* come la combinazione di tante singole parole, per cui il sentimento complessivo dello stesso *tweet* sarà dato dalla somma dei sentimenti delle singole parole, assegnati sulla base di uno specifico dizionario. Esistono diverse tipologie di dizionari che differiscono l'un l'altro a seconda del valore o delle emozioni che essi assegnano a ciascun termine. Nello specifico, l'analisi presente all'interno di tale capitolo è stata effettuata in tre fasi. Le prime due utilizzando due dizionari differenti, dapprima "Bing", attraverso cui estrarre il sentimento globale dei *tweet* e, successivamente, "nrc" per definire le singole emozioni associate ad essi. La terza, invece, prevede l'unione dei metodi usati nelle due fasi precedenti, per delineare dettagliatamente la percezione degli utenti su Twitter relativamente alla *privacy*, rappresentando graficamente sia i *tweet* con più *retweet*, sia i *tweet* con più *like*.

Nella prima fase, quindi, è stato utilizzato il dizionario "Bing" che associa ad ogni termine, in modo binario, un sentimento positivo o negativo. Dei 13.075 *tweet* appartenenti al *dataset* originale, 7.006 sono quelli effettivamente utilizzabili in questo caso, poiché contenenti almeno una delle 6.786 parole presenti al suo interno. Al fine di calcolare il sentimento complessivo di un *tweet*, è stato assegnato valore +1 ai termini classificati positivamente, e valore -1 a quelli classificati negativamente. Sommando i valori delle singole parole si ottiene che il sentimento complessivo di ciascun *tweet* è racchiuso all'interno di un intervallo compreso tra -5 e +7, dove -5 è il punteggio totale del più negativo e +7 invece è quello del più positivo. Di conseguenza è possibile estrarre il sentimento generale dei *tweet*. In particolare, si ottiene che quelli positivi (punteggio > 0) rappresentano il 59,4% del totale, a fronte di quelli negativi (punteggio < 0) che sono il 29,4% e di quelli neutri (punteggio = 0) che sono l'11,2%.

Figura 13 - Pie chart raffigurante il sentimento generale dei tweet



Una maggioranza di *tweet* positivi indica, in generale, che tra gli utenti c'è una buona percezione relativamente alla protezione dei propri dati e della propria *privacy*, quindi essi si sentono tutelati e rassicurati. Il grafico, però, rappresenta anche *tweet* di carattere negativo sebbene in percentuale inferiore. Tale sentimento potrebbe essere suscitato sia dalle notizie sempre più frequenti di attacchi *hacker*, truffe online e furti di identità, sia dall'essere stati vittime di tali reati, sia da un sentimento avverso nei confronti della tecnologia che comunque non si può disconoscere.

La seconda fase dell'analisi, invece, prevede l'utilizzo del dizionario "nrc" che associa ad ogni termine un'emozione. Le emozioni riconosciute da tale dizionario sono 10 di cui due generali (positivo e negativo) e otto specifiche (rabbia, aspettativa, disgusto, paura, gioia, tristezza, sorpresa e fiducia). Associando le parole contenute in ciascun *tweet* con l'emozione corrispondente è stato possibile ottenere l'insieme di stati d'animo che caratterizzano ogni *tweet* e la frequenza con cui essi compaiono al loro interno. Ad esempio, analizzando tramite il dizionario "nrc" i tre *tweet*, definiti precedentemente attraverso il dizionario "Bing", più positivi ed i tre più negativi (Tabella 6), si ottiene che, sia in un caso, sia nell'altro, il sentimento che prevale è quello generico (positivo o negativo) seguito poi da altri più specifici. I *tweet* positivi sono caratterizzati prevalentemente, anche, da un senso di fiducia (*trust*) e di aspettativa (*anticipation*), l'unica nota negativa presente è la paura (*fear*) nel primo *tweet*, ma essendo relativo ad un ambito, come quello medico delle diagnosi e delle cure, è perfettamente comprensibile che esso possa suscitare un po' di timore. I *tweet* negativi, invece, risultano più omogenei dal punto di vista delle emozioni poiché, ve ne sono diverse che appaiono con la stessa frequenza, come per esempio la rabbia (*anger*), la paura (*fear*) e la tristezza (*sadness*).

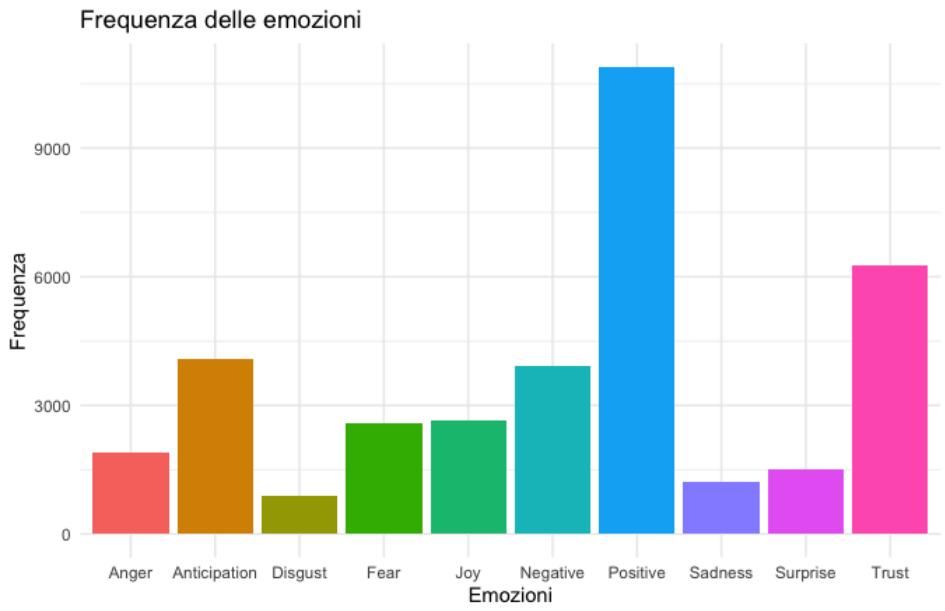
Tabella 5 - Emozioni specifiche dei 3 tweet più positivi e negativi

	ID TWEET	TWEET	EMOZIONI	FREQ. PAROLE
POSITIVI	2241	Dr Vin Menon @DrVinMenon Ensuring that doctors have access to all relevant #medical information. This can lead to more accurate diagnoses and better treatment outcomes. Medical #data is highly sensitive & must be kept secure to protect patient #privacy.	anticipation fear positive trust	3 2 8 4
	513	marjjannnn @marjjannnn Impressed by ZetaChain's focus on privacy and security, utilizing innovative techniques for secure data sharing. Exciting potential for impact in industries like healthcare and finance. #ZetaChain #Privacy #Security #innovation @ZetaBlockchain	anticipation joy positive surprise	1 1 3 1
	9656	Veiovia @veiovia #Privacy can also facilitate greater trust between parties, as users can be confident that their #data is being handled responsibly and transparently. This can lead to more efficient transactions and reduced costs for businesses	anticipation joy positive trust	1 1 5 3
NEGATIVI	2097	Woodhull Freedom Foundation @WoodhullFreedom 2/ These apps scream privacy violation! Article 12 of the Universal Declaration of Human Rights states, "No one shall be subjected to arbitrary interference with his privacy.." #UDHR #humanrights #privacy	anger disgust fear negative sadness surprise	2 1 2 4 2 2
	3012	Faisal Yahya @faisaly #Ring, a home security & smart home company owned by #Amazon, suffers #ransomware attack by Russia-linked ALPHV group. The group threatens to leak stolen data if the company refuses to pay the ransom. #cybersecurity #privacy buff.ly/3ZPaHrW	anger anticipation fear joy negative positive trust	3 1 2 1 4 1 1
	427	ScamNews @ScamNews_ This phishing attack is all over the internet and stealing millions! scam-news.org/2023/03/17/don.. #Scam #scamalert #CrimeNews #scammers #scamwatch #Security #privacy #news #CMC #Crypto #Binance #ElonMuskNews #Tesla #phishing #stocks #fraud #SCAM #scamming #CrimeNews	anger disgust fear negative	2 1 2 3

Per quanto riguarda gli strumenti utilizzati per effettuare tale analisi è doveroso affermare che, i risultati che si ottengono non sono del tutto ragionevoli, poiché ottenuti tramite processi automatizzati che restituiscono il sentimento generale di un *tweet* a partire dalla somma dei singoli termini. Inoltre, non tengono in considerazione aspetti cruciali quali l'ironia, il ruolo che una parola ricopre all'interno di un periodo e l'impatto del contesto esterno sull'argomento in questione.

Attraverso il grafico sottostante, che riporta per ogni emozione il numero totale di parole che compongono i *tweet* ad essa associata, è possibile ottenere un'immagine generale degli stati d'animo caratterizzanti il *dataset* preso in analisi. Da esso si evince una prevalenza del sentimento positivo rispetto al negativo, giustificata dal fatto che, come detto in precedenza, gli individui si sentono maggiormente protetti e tutelati da aziende ed istituzioni. Questo li porta a sviluppare un forte senso di fiducia nei loro confronti, e conseguentemente di entusiasmo ed aspettativa per gli sviluppi futuri. A fronte di tale prospettiva positiva, è normale che essi provino gioia, ma allo stesso tempo anche timore perché molto spesso, ci si aspetta l'imprevisto non pianificabile. Valori elevati di fiducia, inoltre, giustificano un basso livello di sentimenti legati alla sorpresa perché qualsiasi cosa possa accadere si ha anche fiducia che possa essere risolta positivamente.

Figura 14 - Grafico raffigurante il numero di parole che compongono i tweet associate ad ogni emozione



Analizzare i sentimenti e le emozioni dei propri consumatori aiuta a comprendere che impatto abbiano determinati argomenti, comportamenti o prodotti su di essi. L'analisi del sentimento, infatti, è uno strumento molto utilizzato nel mondo del marketing proprio in virtù dei numerosi vantaggi che esso offre, tra cui la possibilità di migliorare le relazioni con i clienti, la reputazione aziendale e l'efficacia delle strategie di marketing.

La terza fase, infine, prevede l'unione, attraverso una rappresentazione grafica, delle informazioni ottenute dall'analisi dei *tweet* con più *retweet* e con più *like*, con i risultati della *sentiment analysis*, effettuata nelle due fasi precedenti. I *tweet* con più *retweet* e con più *like*, infatti, riguardano solitamente contenuti che hanno suscitato maggiore interesse e

coinvolgimento da parte degli utenti. Analizzandoli, quindi, è possibile ottenere una visione più approfondita delle emozioni e dei sentimenti degli utenti relativi ad argomenti e informazioni rilevanti per il tema della *privacy* e della protezione dei dati personali. Nello specifico, i grafici realizzati sono due: uno relativo ai *tweet* con il maggior numero di *retweet* e l'altro, invece, ai *tweet* con il maggior numero di *like*. Entrambi presentano sull'asse delle ascisse il sentimento generale ottenuto utilizzando il dizionario "Bing" e, quindi, compreso all'interno dell'intervallo (-5, +7), mentre sull'asse Y il valore corrispondente al numero di *like* e di *retweet*. Per migliorare la rappresentazione grafica, facilitandone anche l'interpretazione, i valori dell'asse Y sono stati raffigurati in scala logaritmica, usando log(1+ retweet_count) per i *retweet* e log(1+ favorite_count) per i *like*. Un ulteriore elemento importante è legato ai punti. Essi, infatti, presentano forme differenti a seconda del sentimento a cui sono associati grazie al dizionario "nrc", tuttavia è possibile che ad un *tweet* siano associati più sentimenti differenti, qualora essi appaiano al suo interno con eguale frequenza massima. Tale è il motivo per cui all'interno del grafico alcune forme risultano sovrapposte.

Figura 15 – Rappresentazione grafica dei retweet in funzione dei sentimenti ottenuti tramite dizionario 'Bing' e 'nrc'

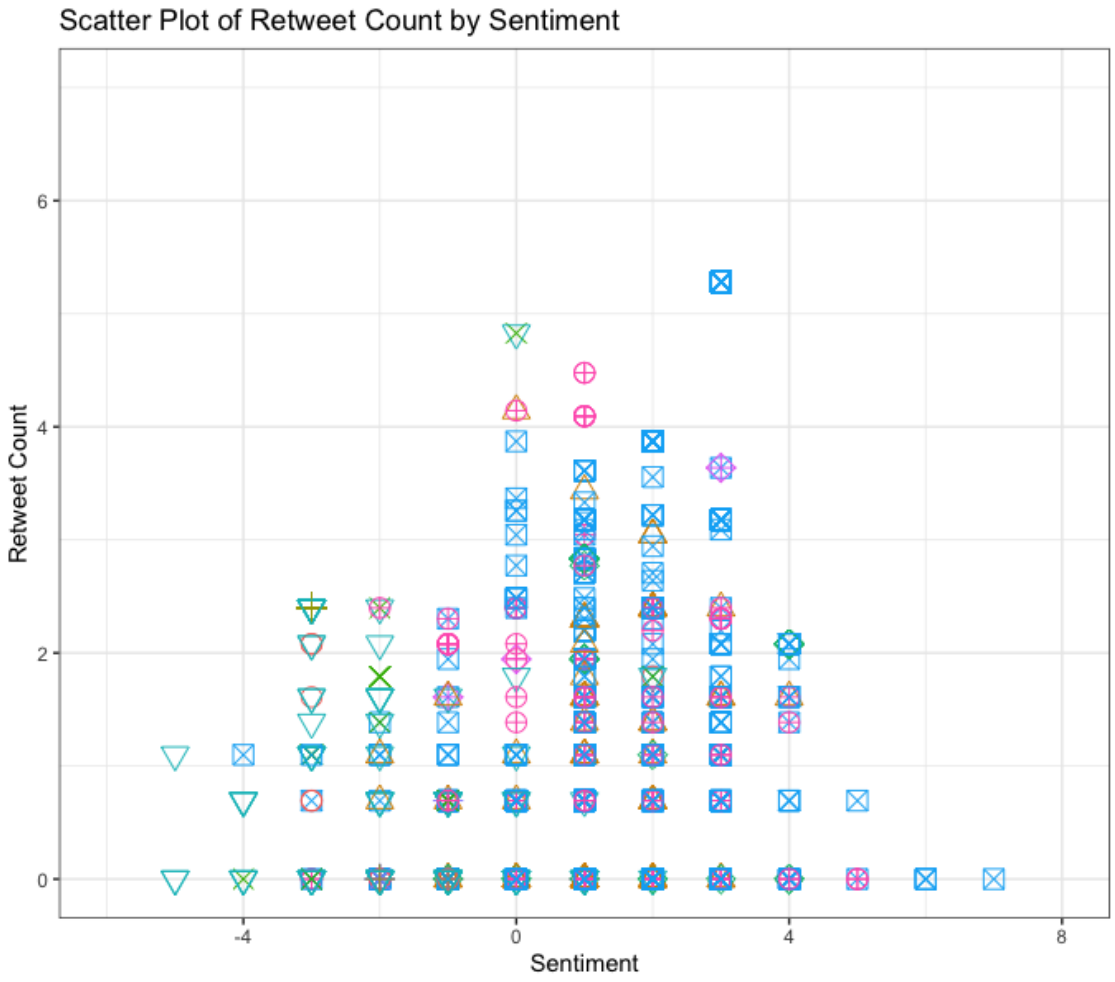
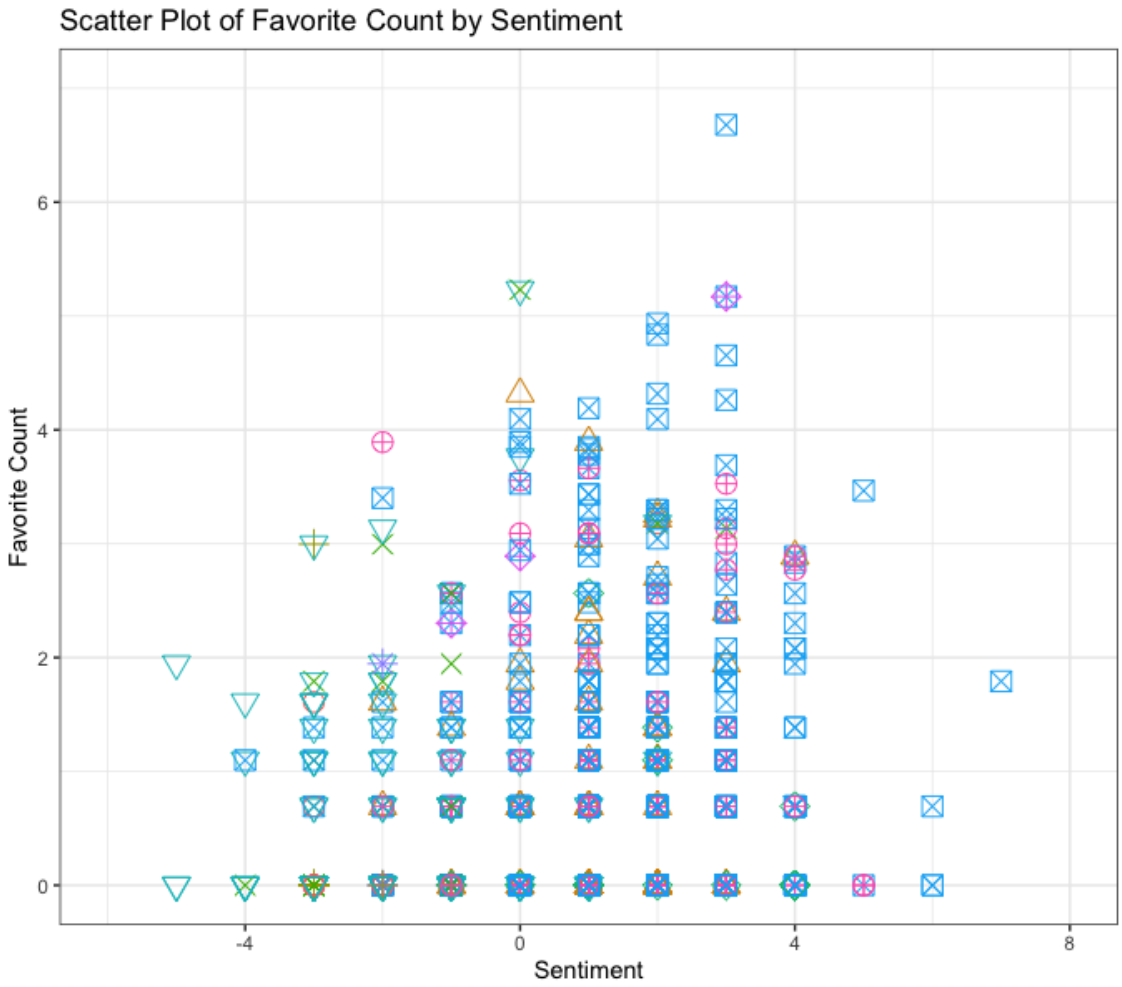


Figura 16 - Rappresentazione grafica dei like in funzione dei sentimenti ottenuti tramite dizionario 'Bing' e 'nrc'



Prendendo in considerazione il grafico relativo ai *retweet*, i *tweet* con il maggior numero di retweet sono principalmente localizzati nell'intervallo tra 0 e 2. Tali *tweet*, inoltre, sono caratterizzati da un sentimento prevalentemente positivo, seguito da emozioni come la fiducia, la sorpresa e l'aspettativa. Allo stesso modo, nel grafico relativo ai *like* i *tweet* con il maggior numero di *like*, sono localizzati prevalentemente nell'intervallo tra 1 e 3, e sono caratterizzati da sentimenti positivi di fiducia e di aspettativa.

Sulla base di quanto detto fino a qui, è possibile effettuare una riflessione relativamente alla disposizione dei punti nei due grafici. Il fatto che in entrambi i casi il maggior numero di *tweet* con più *retweet* e di *tweet* con più *like* si trovi al centro della distribuzione, per cui in posizione neutrale rispetto alla polarità generale del sentimento, potrebbe da un lato indicare che gli individui sono generalmente soddisfatti rispetto alle forme di tutela della loro *privacy* e alle misure di protezione adottate, ma allo stesso tempo potrebbe indicare che gli utenti si ritengono

indifferenti rispetto alla questione della *privacy*, esprimendo un'opinione bilanciata tra sentimento positivo e negativo, piuttosto che esporsi apertamente.

3.4 – RISULTATI

A conclusione dell'analisi eseguita, volta ad indagare la percezione degli utenti relativamente alla tutela della propria *privacy* e dei propri dati personali, è doveroso presentare i risultati ottenuti. Dall'analisi esplorativa condotta sulle parole più frequenti, sugli *hashtag* più usati, sui *tweet* con più *retweet* e sui *tweet* con più *like*, è emerso che su Twitter il dibattito relativo al tema della *privacy* si articola attorno ad elementi molto attuali e quotidiani, quali la sicurezza informatica, l'utilizzo improprio dei dati personali, l'uso considerevole di *social network* e di piattaforme digitali e, soprattutto, l'ambito degli investimenti finanziari legati alla *blockchain* ed alle criptovalute. La centralità di quest'ultimo tema legato al dibattito sulla *privacy* dimostra consapevolezza e riconoscimento da parte degli utenti in merito alle caratteristiche di sicurezza e *privacy* offerte da tali tecnologie. Al riguardo, infatti, l'informatizzazione dei servizi bancari ha condotto ad un uso sempre più massivo delle piattaforme digitali per le operazioni da effettuare sui propri risparmi e sulla gestione in generale delle proprie finanze. È chiaro che la tutela del dato personale, in questo caso quello economico, legato alla propria storia personale, lavorativa e familiare, assume una valenza molto alta e, proprio in virtù di ciò, un valore fondante nel rapporto con l'interlocutore commerciale/bancario/istituzionale.

L'analisi del sentimento restituisce un sentimento generale del *dataset* preso in esame di carattere positivo, con emozioni prevalenti come la fiducia e l'aspettativa. Ciò vale sia per i *tweet* con il maggior numero di *retweet*, relativi perlopiù al tema della *blockchain*/criptovalute, sia per i *tweet* con più *like*, relativi all'aspetto della sicurezza e della protezione dei dati.

La percezione positiva legata a temi attuali indica, quindi, maggior fiducia da parte degli individui relativamente alla tutela della propria identità e dei propri dati. Ed il dato finale di sintesi è importante perché si è visto come in un dibattito online fatto di brevi messaggi scambiati virtualmente, l'aspetto della tutela della *privacy*, ovvero della propria identità personale, si unisca a quello più materiale del lato economico della vita, ricercando la sicurezza di quei dati che ne delineano i contorni e ne definiscono il peso monetario.

In ogni caso è importante tenere a mente che, essendo la *privacy* un tema così attuale ed in costante evoluzione, i risultati ottenuti rispecchiano la percezione della società attuale che differisce, però, da quella che avevano, relativamente ad essa, gli individui due decenni fa e da quella che gli stessi avranno fra vent'anni. Analizzare e comprendere, quindi, una tematica come quella relativa alla tutela della propria identità e dei propri dati personali è un compito

che merita di essere eseguito costantemente, considerando anche l'impatto che essa ha nel quotidiano.

CONCLUSIONI

"Le emozioni guidano le decisioni di acquisto. Conoscere queste emozioni è fondamentale per avere successo nel marketing." (Daniel Kahneman)

Alla luce di quanto esaminato nel presente elaborato, emerge chiaramente l'importanza della *privacy* e della protezione dei dati personali nella società digitale odierna. Si può, infatti, affermare che esse rappresentino due sfide significative per una realtà sempre più connessa, come quella in cui viviamo. Tali concetti, per quanto dibattuti e diffusi, spesso vengono erroneamente considerati sinonimi, creando confusione e malintesi. La *privacy*, in realtà, riguarda il diritto di un individuo a mantenere il controllo e la riservatezza delle proprie informazioni personali, decidendo quali dati condividere, con chi ed in che modo, tutelando, così, la propria identità. D'altro canto, invece, la protezione dei dati personali riguarda le misure adottate da organizzazioni, aziende ed istituzioni per garantire la sicurezza e l'integrità dei dati e delle informazioni che vengono raccolte e trattate.

Gran parte dei timori e delle preoccupazioni attuali relative ai due concetti sopra citati, derivano dalla raccolta e dall'utilizzo massiccio di dati provenienti da *social media*, denominati anche *social media data*. Quest'ultimi, vengono ampiamente utilizzati nel mondo del marketing poiché consentono alle aziende di creare strategie personalizzate, offrire prodotti e servizi mirati alle preferenze degli individui, migliorando l'esperienza complessiva del consumatore. Tuttavia, è fondamentale che le stesse operino sempre in modo etico e rispettoso della *privacy* e della sfera personale degli individui. Un ulteriore utilizzo dei *social media data*, nel campo del marketing, è rappresentato dall'analisi del sentimento. Questo approccio si avvale di dati ed informazioni dei clienti al fine di comprendere la loro percezione emozionale relativamente a prodotti, brand, eventi ed argomenti con l'obiettivo di garantire loro piena soddisfazione.

Al riguardo, dall'analisi effettuata nel terzo capitolo attraverso lo strumento della *sentiment analysis* è emerso che gli individui hanno, tendenzialmente, una percezione positiva della tutela della propria *privacy* e dei propri dati personali, legata soprattutto a temi attuali quali social network, piattaforme digitali, *blockchain*, criptovalute ed attacchi *hacker*. Comprendere la percezione degli individui aiuta le aziende a definire strategie appropriate, che sappiano essere efficaci ed efficienti. Nel caso della *privacy*, per esempio, ad un'azienda consapevole di avere clienti molto attenti e preoccupati per la propria identità digitale, converrà adottare una strategia

che punti a comunicare il suo impegno da questo punto di vista, cercando di rassicurare il cliente e dimostrando di potergli offrire tutto quello di cui ha bisogno. Aziende che sono in grado di dimostrare ai clienti di avere a cuore tale aspetto di protezione e di tutela, riescono infatti a suscitare in loro quel senso di fiducia e di tranquillità fondamentale per la creazione di un rapporto solido e duraturo, basato su collaborazione, rispetto e fedeltà. Ciò contribuisce, quindi, a rendere gli utenti maggiormente disponibili al rilascio di dati ed informazioni personali, così da ricevere offerte e contenuti personalizzati. Al contrario, invece, un cliente che manifesta una percezione negativa in merito alla tutela dei propri dati ed informazioni non ha fiducia nell'azienda, quindi non si dimostrerà disposto ad instaurare legami, a collaborare ed a fornire dati. In tal caso, se l'azienda ne ha possibilità, vale la pena provare a fargli cambiare idea, altrimenti è importante capire, in modo costruttivo, da cosa derivi tale percezione, così da modificare o risolvere eventuali problemi e mancanze.

In sintesi, in un panorama digitale in continua evoluzione, la tutela della *privacy* e dei dati personali rimane un aspetto cruciale per garantire la fiducia degli utenti e il rispetto dei loro diritti fondamentali.

BIBLIOGRAFIA

H. ARENDT, *Vita Activa. La condizione umana*, trad. it. Di A. Dal Lago, 2001, p.9.

ARISTOTELE, *La Politica*, Laterza, 2007.

A. ACQUISTI, *Privacy in Electronic Commerce and the Economics of Immediate Gratification,* Proceedings of the 5th ACM Conference on Electronic Commerce, 2004, pp. 21-29.

L. ARTHUR, Big Data Marketing: Engage Your Customers More Effectively and Drive Value, Wiley, 2013.

L. BRANDIMARTE, A. ACQUISTI, G. LOEWENSTEIN, *Misplaced Confidences: Privacy and the Control Paradox*, *Social Psychological and Personality Science* 4 (3), 2013, pp. 340-347.

G. BUTTARELLI, *Banche dati e tutela della riservatezza*, Giuffrè editore, 1997.

F. CARNELUTTI, *Diritto alla vita privata*, Giuffrè, Milano, 1955, cit. p. 3.

M. CASTELLS, *Communication Power,* Oxford University Press, 2009.

A. CERON, L. CURINI, M. IACUS, *Social Media e Sentiment Analysis, l'evoluzione dei fenomeni sociali attraverso la Rete*, Springer, 2013.

M. CHUI, M. COLLINS, M. PATEL, *The Internet of Things: Catching up to an accelerating opportunity,* McKinsey Global Institute, 2021.

A.C. DI LANDRO, *Big Data. Rischi e tutele nel trattamento dei dati personali*, Edizioni Scientifiche Italiane, 2020, p.40.

M. FERRARA SANTAMARIA, *Il diritto alla illesa intimità privata*, *Riv. dir. civ*., 1937, I, cit. p. 168.

G. FIORIGLIO, *Il diritto alla privacy. Nuove frontiere nell'era di internet*, 2008, pp. 9-10.

L. FLORIDI, M. TADDEO, *What is Data Ethics*, *Philosophical transactions of the Royal Society of London*. Series A: Mathematical, physical and engineering sciences, 2016, Vol.374 (2083), p.20160360.

A. U. FOX, *Social Media Analytics Strategy. Using Data to Optimize Business Performance*. Apress, 2022.

G. GREENWALD, *No place to hide – Sotto controllo: Edward Snowden e la sorveglianza di massa,* Rizzoli, Milano, 2014.

M. HU, *Small Data surveillance v. Big Data cybersurveillance*, in 42 Pepp. L. Rev. 773, 2015.

M. IASELLI, S. GORLA, *Storia della Privacy*, Lex Et Ars, 2015.

INTERNET SOCIETY, *The Internet of Things: An Overview. Understanding the Issues and Challenges of a more Connected World*, 2015, p.21.

M. JAMES, M. CHUI, P. BISSON, J. WOETZEL, R. DOBBS, J. BUGHIN, D. AHARON, *Mapping the world beyond the hype*, McKinsey Global Institute, 2015.

A. KAPLAN, M. HAENLEIN. *Users of the World: Unite! The Challenges and Opportunities of Social Media*. Business Horizons, 2010.

S. KOKOLAKIS, *Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon*, Computers & Security 64, 2017, pp. 122 – 134

P. KOTLER, G. ARMSTRONG, F. ANCARANI, M. COSTABILE, *Principi di Marketing*, Pearson, 2019.

K. KOYCE, *The Challenges of Using Big Data Effectively. A critical analysis of the phenomenon of big data through the parameters of the end-user, industry uses and legal considerations*, 6th Annual International Conference on Enterprise Marketing and Globalization (EMG), 2017.

F. LIGI, *Il diritto alle vicende e la sfera della personalità, Foro italiano*, Roma, 1955, I, cit. p. 386.

N. LUGARESI, *Internet, Privacy e Pubblici Poteri negli Stati Uniti*, Dott. A. Giuffrè Editore, 2000, p.223.

A. MANDELLI, *Big Data Marketing. Creare valore nella platform economy con dati, intelligenza artificiale e IoT*, 2017, EGEA edizioni.

A. MANTELERO, *Il costo della privacy tra valore della persona e ragione d'impresa*, Giuffrè, Milano, 2007, cit. p.1.

K. D. MARTIN, P. E. MURPHY, *The role of data privacy in marketing*, Journal of the Acad. Mar. Sci. 45, 2017, pp. 135-155

M. MENSI, P. FALLETTA, *Il diritto del web. Casi e materiali*, Cedam, 2015, cap. XI, p. 329.

L. MIGLIETTI, *Profili Storico-Comparativi del Diritto alla Privacy*, 2014, pp 3-4.

S. NIGER, *Le nuove dimensioni della privacy: dal diritto alla riservatezza alla protezione dei dati personali*, 2006

A. OLTEANU, C. CASTILLO, F. DIAZ, E. KICIMAN, *Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries*, 2019, Front. Big Data 2:13.

H. NISSENBAUM, *A Contextual Approach to Privacy Online*, 2011, pp. 32-45.

U. PAGALLO, *La tutela della privacy negli Stati Uniti d'America e in Europa*, Giuffrè, Milano, 2008, p. 96 e ss.

(a cura di) R. PANNETTA, *Libera circolazione e protezione dei dati personali*, Milano, Giuffrè, 2006, v. nota 2, tomo I, Parte Prima, SABINA KIRSHEN, *Il codice della privacy tra tradizione ed innovazione*, p. 80.

PEW RESEARCH CENTER, *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*, 2019.

F. PIZZETTI, *Datagate, Prism, caso Sonwden: il mondo tra nuova grande guerra cibernetica e controllo globale*, in Federalismi, 26 Giugno 2013.

F. PIZZETTI (a), *La protezione dei dati personali dalla direttiva al nuovo regolamento*, pp. 55-57 in G. Busia, L. Liguori, O. Pollicino (a cura di), *Le nuove frontiere della privacy nelle tecnologie digitali. Bilanci e prospettive*, Aracne editrice, 2016.

M. F. PYNADATH, T. M. ROFIN, S. THOMAS. *Evolution of customer relationship management to data mining-based customer relationship management: a scientometric analysis*, Quality & Quantity, Springer, 2022.

F. PIZZETTI (b), *Privacy e il diritto europeo alla protezione dei dati personali. Il Regolamento europeo 2016/679*, vol. II, Giappichelli Editore, 2016.

G. RESTA, *La sorveglianza elettronica di massa e il conflitto regolatorio USA/UE*, in Diritto dell'Informazione e dell'Informatica, 2015, pp. 697-718.

A. RICHTERICH, *The Big Data Agenda: Data Ethics and Critical Data Studies*, Univeristy of Westminster Press, 2018.

S. RODOTÀ, *Discorso di presentazione della relazione annuale del garante al parlamento*, 2001.

S. RODOTÀ, *Il diritto di avere diritti*, Laterza, Bari, 2015, p.320.

S. RODOTÀ, *La privacy tra individuo e collettività*, Il Mulino, 1974.

S. RODOTÀ, *Tecnologie e diritti*, Bologna, il Mulino, 1999; A. WESTIN, *op.cit.*.

M. SOFFIENTINI, *Accountability e figure privacy: Titolare, Responsabile e Incaricato e DPO*, 2017. Diritto e pratica del lavoro 39, pp. 2333-2338.

D. SOLOVE, *The Meaning and Value of Privacy*, in D. Mokrosinska (a cura di), *Social Dimensions of Privacy: Interdisciplinary Perspectives*, 2015, pp. 71-82.

L. TEMPESTINI, G. D'ACQUISTO, *Il dato personale oggi tra le sfide dell'anonimizzazione e le tutele rafforzate dei dati sensibili*, pp. 79-80 in G. Busia, L. Liguori, O. Pollicino (a cura di), *Le nuove frontiere della privacy nelle tecnologie digitali. Bilanci e prospettive*, Aracne editrice, 2016.

S.D. WARREN, L.D. BRANDEIS, *The Right to Privacy*, in *Harvard Law Review*, 1890, 5, pp. 193-220.

We Are Social. *Global Digital Statistics 2023*, 2023.

M. WEDEL, P. K. KANNAN, *Marketing Analytics for Data- Rich Environments*, Journal of Marketing, 80, 2016.

A. WESTIN, *Privacy and Freedom*, New York, 1967.

Y. XU, D. C. YEN, B. LIN, D. C. CHOU, *Adopting customer relationship management technology*, Industrial Management & Data Systems, 2002, Vol. 102 No. 8, pp. 442-452.

G. ZICCARDI, *I flussi d'informazione digitale e la loro sicurezza nel panorama geopolitico attuale*, 2014.

S. ZUBOFF, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, 2019.

A. ZWITTER, *Big Data Ethics*, Big Data & Society vol.1, 2014.

SITOGRAFIA

Agenda Digitale, 2020. *Etica e trattamento dati: così si sconfigge la dittatura dell'algoritmo*. Disponibile su <https://www.agendadigitale.eu/sicurezza/privacy/etica-e-trattamento-dati-cosi-si-sconfigge-la-dittatura-dellalgoritmo/> (Ultima consultazione: 27 aprile)

Agenda Digitale (a), 2019. *GDPR, che si intende per dati personali: natura, tipologie e qualità dei dati sensibili*. Disponibile su < https://www.agendadigitale.eu/cittadinanza-digitale/gdpr-che-si-intende-per-dati-personali-natura-tipologie-e-qualita/ > (Ultima consultazione: 20 febbraio)

Agenda Digitale (b), 2019. *Privacy vs protezione dati personali: attenti alla differenza, ne va della nostra identità.*
Disponibile su <https://www.agendadigitale.eu/sicurezza/privacy/privacy-e-protezione-dati-personali-cosa-sono-quali-differenze-cosa-e-cambiato-col-gdpr/#California_Consumer_Privacy_Act_e_Gdpr_a_confronto > (Ultima consultazione: 15 febbraio)

Analisi Difesa, 2023. *Polizia Postale: attacchi informatici in crescita del 138% nel 2022.* Disponibile su <https://www.analisidifesa.it/2023/01/polizia-postale-attacchi-informatici-in-crescita-del-138-nel-2022/ > (Ultima consultazione: 2 marzo)

Commissariato di P.S. online, 2023. *Resoconto attività 2022 della polizia postale e delle comunicazioni e dei centri operativi sicurezza cibernetica.*
Disponiblie su < https://www.commissariatodips.it/notizie/articolo/resoconto-attivita-2022-della-polizia-postale-e-delle-comunicazioni-e-dei-centri-operativi-sicurezz/index.html > (Ultima consultazione: 27 febbraio)

Corcom, 2022. *Privacy, Onu: "Sempre più rischi nell'era digitale, urgente la regulation".* Disponibile su < https://www.corrierecomunicazioni.it/privacy/privacy-onu-sempre-piu-rischi-nellera-digitale-urgente-la-regulation/ > (Ultima consultazione: 1 marzo)

CyberSecurity 360, 2022. *Attacchi hacker: strumenti e tecniche dei nuovi cyber criminali.* Disponibile su <https://www.cybersecurity360.it/nuove-minacce/attacchi-hacker-strumenti-e-tecniche-dei-nuovi-cyber-criminali/ > (Ultima consultazione: 4 marzo)

CyberSecurity 360, 2020. Cookie: cosa sono, a cosa servono e quali regole privacy seguire. Disponibile su https://www.cybersecurity360.it/legal/privacy-dati-personali/cookie-cosa-sono-a-cosa-servono-e-quali-regole-privacy-seguire/ (Ultima consultazione: 3 maggio)

Garante per la protezione dei dati personali. *Cosa intendiamo per dati personali?*
Disponibile su < https://www.garanteprivacy.it/home/diritti/cosa-intendiamo-per-dati-personali > (Ultima consultazione: 17 febbraio)

General Networking, 2020. *Cosa vuol dire segmentare e perché è importante per l'azienda.* Disponibile su <https://www.gnet.it/blog-gn/cosa-vuol-dire-segmentare-e-perché-è-importante-per-l-azienda> (Ultima consultazione: 4 maggio)

Il Sole24Ore, 2023. *Cybersecurity, 2022 annus horribilis: 13mila attacchi, +138%.* Disponibile su <https://www.ilsole24ore.com/art/cybersecurity-2022-annus-horribilis-13mila-attacchi-138percento-AE2W4pTC?refresh_ce> (Ultima consultazione: 2 marzo)

Informatica e Ingegneria online, 2021. *Differenza tra marketing tradizionale e marketing digitale*. Disponibile su https://vitolavecchia.altervista.org/la-differenza-tra-marketing-tradizionale-e-marketing-digitale/ (Ultima consultazione: 8 marzo)

Inside marketing (a), 2019. *Big Data cosa sono e perché sono importanti per le aziende.* Disponibile su <https://www.insidemarketing.it/glossario/definizione/big-data/#le-5v-dei-big-data> (Ultima consultazione: 30 marzo)

Investopedia, 2020. *Social Data: What is it, How it Works, Limitations.* Disponibile su <https://www.investopedia.com/terms/s/social-data.asp > (Ultima consultazione: 6 giugno)

Juicer, 2021. *Your Social Media Data: What's Collected and How is it Used?* Disponibile su <https://www.juicer.io/blog/your-social-media-data-what-s-collected-and-how-is-it-used > (Ultima consultazione: 6 giugno)

Namirial Focus, 2022. *Furto identità digitale: come avviene e come difendersi.* Disponibile su <https://focus.namirial.it/furto-identita-digitale/> (Ultima consultazione: 1 marzo)

Privacy Lab, 2020. *GDPR e marketing: la profilazione.* Disponibile su https://www.privacylab.it/IT/1042/gdpr-e-marketing-la-profilazione/#:~:text=Perché%20la%20maggior%20parte%20dei,non%20riguarda%20solo%20l%27online (Ultima consultazione: 1 maggio)

Statista, 2021. *Volume of data/information created, captured, copied and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025*. Disponibile su <https://www.statista.com/statistics/871513/worldwide-data-created/ > (Ultima consultazione: 27 marzo)

Tech Enthusiast, 2019. *Quanti dati su Internet viaggiano in un minuto? Le risposte.* Disponibile su < https://www.techenthusiast.it/quanti-dati-su-internet-viaggiano-in-un-minuto/> (Ultima consultazione: 28 marzo)

The Washington Post, 2013. *U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program.*

Disponibile su <https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html > (Ultima consultazione: 12 marzo)

Trend Online, 2021. Digital marketing: differenze con il marketing tradizionale. Disponibile su <https://www.trend-online.com/tecnologia/digital-marketing-tradizionale-marketing/ > (Ultima consultazione: 3 maggio)

World Federation of Advertisers (WFA), 2020. *Data Ethics – The Rise of Morality in Technology*. Disponibile su <www.wfanet.org/dataethics> (Ultima consultazione: 27 aprile)

World Federation of Advertisers (WFA), 2022. *Data ethics is a priority for nine out of 10 CMOs but half need help making it a reality, WFA research shows.* Disponibile su < https://wfanet.org/knowledge/item/2022/09/28/Data-ethics-is-a-priority-for-nine-out-of-10-CMOs-but-half-need-help-making-it-a-reality-WFA-research-shows> (Ultima consultazione 27 aprile)

RINGRAZIAMENTI

Innanzitutto ci tengo a ringraziare il Professor Aliverti, che ha saputo essere un relatore sempre positivo, presente ed attento, attraverso i suoi numerosi spunti e consigli preziosi per la realizzazione di questo elaborato.

Grazie alla mia famiglia e a Francesco per non aver mai smesso di credere in me, motivandomi e spronandomi a dare sempre il meglio.

Un grazie a Soemi, Nicola e a tutti quei compagni di corso con cui ho condiviso le gioie e le fatiche dei lavori di gruppo, delle lezioni e degli esami, senza di voi sarebbe stato tutto un po' più noioso.

Ringrazio anche gli amici e le amiche di sempre per aver reso tutti i momenti di pausa dallo studio divertenti e spensierati, vi voglio bene!

Infine, vorrei rivolgere un grazie anche a me stessa, per la determinazione e forza di volontà con cui affronto ogni sfida, mettendomi sempre in gioco e cercando di raggiungere, in un modo o nell'altro, tutti gli obiettivi che mi pongo.

Ora più che mai, quello che mi dice sempre la nonna è proprio vero: *"ad maiora, semper!"*