Sommario

INTRODUZIONE

Con l'avvento dell'era digitale è diventato consuetudine condividere regolarmente i propri dati personali su internet per registrarsi su innumerevoli piattaforme e beneficiare dei loro molteplici servizi. Questi dati che gli utenti lasciano online hanno assunto un ruolo molto importante poiché contengono informazioni sensibili su individui (nome, cognome, indirizzo, informazioni finanziarie, dati di localizzazione e altro) le quali vengono continuamente generate, raccolte, rielaborate e condivise in rete. Un esempio lampante di questo fenomeno è rappresentato dalle grandi piattaforme digitali, definite come "delle architetture programmabili progettate per organizzare le interazioni tra utenti" (Carmen Pupo, 12 aprile 2023), le quali hanno gradualmente consolidato un'enorme fetta di potere negli ultimi anni. Ogni volta che un utente visita una pagina online per visualizzare un post pubblicitario relativo a un articolo che ha precedentemente cercato, la piattaforma gli chiede di accettare i propri termini e condizioni riguardo al trattamento dei dati personali. In pratica, la piattaforma richiede il consenso per monitorare le sue ricerche e preferenze, al fine di proporre contenuti pubblicitari simili in futuro, personalizzando l'esperienza dell'utente. Infatti, le piattaforme hanno la capacità di trasformare in dati tutte le interazioni utente-piattaforma (mettere un like, fare un retweet, richiedere amicizie), monitorando i loro interessi con l'obbiettivo di predire i loro potenziali acquisti futuri. Da un lato, queste piattaforme rendono possibili e facilitano le transazioni tra inserzionisti e consumatori, creando opportunità che altrimenti non sarebbero disponibili. Dall'altro, però, sollevano preoccupazioni relative alla privacy e alla sicurezza degli utenti. Le informazioni che gli utenti condividono online, infatti, spesso non sono facilmente cancellabili una volta che decidono di non volerle più divulgare, e tendono a rimanere accessibili sulla rete per un periodo indeterminato. Nonostante in molti casi l'utilizzo di questi dati è per scopi legittimi, ad esempio miglioramento di un servizio, personalizzazione dell'esperienza del cliente o condurre analisi sul mercato, non scompare il rischio che alcuni soggetti terzi possano, invece, usufruirne per scopi illeciti che mettono a rischio la privacy e la sicurezza degli individui. Si tratta delle Big Tech, ovvero quei colossi tecnologici definiti dalla Commissione europea come le organizzazioni più influenti nel contesto delle operazioni di mercato, poiché esercitano un notevole impatto nelle relazioni tra consumatori e imprese. Sorge quindi la necessità di indagare se tali piattaforme operino in modo etico nell'utilizzo dei dati dei propri utenti registrati, oppure se sfruttino tali informazioni al fine di ottenere un vantaggio competitivo nel mercato di riferimento.

Il presente elaborato si propone di esaminare nello specifico il caso tedesco di Facebook che è stato accusato dall'Antitrust di abuso della sua posizione dominante, in particolare di combinare i dati degli utenti provenienti da diverse fonti per la creazione di "super profili". Tali sospetti diventano leciti dopo che le piattaforme, con l'inevitabile successo del web, siano passate dal loro originale ruolo di mettere in comunicazione gli utenti ad architetture tramite le quali si creano delle relazioni tra istituzioni e aziende, dove il dato assume un elevato valore commerciale. Se da un lato può sembrare che internet abbia migliorato la nostra vita aumentando il valore percepito dal cliente sull'efficienza del servizio delle piattaforme, dall'altro lato ci si rende conto che fa spesso uso di metodi che mettono a rischio la privacy e la sicurezza degli individui.

Il caso di Facebook coinvolge diversi punti focali di grande attualità sui quali è stata aperta una lunga discussione negli anni recenti: innanzitutto ha messo in luce per la prima volta alcuni aspetti critici sulle Big Tech che hanno reso necessaria l'applicazione di leggi per regolamentare il mercato digitale; in secondo luogo oltrepassa i confini della semplice tutela della concorrenza, cercando di delineare i confini tra le limitazioni economiche e il diritto alla privacy, creando anche un'intersezione tra questi due. L'obiettivo che si propone il caso è quello di determinare se la tutela dei dati personali possa fungere da strumento cruciale per valutare l'esistenza di condizioni di concorrenza effettiva all'interno del mercato. In altre parole, si mira a verificare se le imprese, attraverso il loro posizionamento privilegiato, possano compromettere le condizioni per una sana competizione sfruttando a proprio vantaggio la posizione acquisita.

CAPITOLO 1 – NORMATIVE SULLA PRIVACY

1.1 La nascita di Internet

Precedentemente allo sviluppo della tecnologia, quindi alla quotidiana e massiccia condivisione di dati e informazioni nel web, il concetto di privacy era molto diverso da quello che si è poi sviluppato e che ancora oggi viene continuamente rivisitato. Nel passato, infatti, il diritto alla privacy coincideva con la concezione statunitense del "diritto di essere lasciato solo" e godere di una sfera riservata senza l'intrusione altrui. Il cammino di evoluzione in campo di protezione della privacy fu molto lento e tortuoso in quanto gran parte della società, tra cui alcuni giudici della Corte Suprema, impiegò tempo per comprendere l'importanza e la necessità del riconoscimento di questo diritto. A lungo la natura del concetto fu molto oscillante fino a quando negli anni 60 la giurisprudenza statunitense, con una serie di sentenze, riconobbe la privacy come un oggetto da proteggere sia nella sfera pubblica che privata dell'individuo. Tuttavia, è importante sottolineare che negli Stati Uniti la protezione dei dati viene contestualizzata al solo ambito economico e non viene considerata come un diritto fondamentale del cittadino, ma piuttosto come un diritto del consumatore da bilanciare con l'interesse delle imprese. Anche in Europa la questione sulla protezione del diritto alla privacy ha subito un lungo processo di riconoscimento a causa della forte dipendenza dai mutamenti sociali, culturali e tecnologici.

Nella società in cui viviamo oggi, con l'avvento dell'era digitale, si avverte la costante sensazione che i propri dati personali siano sempre in pericolo e non ci si sente al sicuro nella propria sfera privata, temendo di non avere abbastanza controllo sulle informazioni che vengono condivise online. Si vuole evitare che la gestione sulla circolazione dei dati e delle informazioni riservate delle persone nel web sia gestito da un numero limitato di individui i quali avrebbero così il potere di influenzare la vita degli altri attraverso gli strumenti digitali. Il ruolo dei dati personali nell'era digitale mette in luce la loro crescente rilevanza economica; infatti, costituiscono per le aziende degli asset fondamentali per creare sempre più valore che gli permette di aumentare il loro profitto. Ciò è consentito dalle peculiarità che il dato in sé contiene:

- Il costo marginale di incremento o duplicazione del dato è nullo;
- Il valore di un dato non è unico, ma dipende dalla mole e dalla significatività di informazioni che è in grado di generare;
- L'affidabilità e la potenzialità di un dato sono legati alla sua capacità di poter fare un'attenta analisi di mercato.

Ogni volta che un utente utilizza un dispositivo connesso ad una rete viene generata una grande quantità di dati che le piattaforme possono tracciare per classificare i vari utenti in base ai loro

gusti e interessi personali, creando dei target di mercato. Dopo di che, saranno in grado di generare pubblicità mirata per ogni gruppo di utenti identificato, in modo tale da offrirgli in pochissimo tempo ciò che loro stanno cercando. Tuttavia, questo processo crea un grande dibattito inerente alla violazione dei diritti sulla privacy poiché i dati, una volta generati, possono rimanere disponibili molto più a lungo rispetto alla vita della persona che li ha creati. Inoltre, grazie alla loro caratteristica di poter essere duplicati a costo nullo, le piattaforme possono scambiarli e venderli ad altre aziende senza che gli utenti ne siano a conoscenza. Sarebbe molto più semplice se un individuo potesse eliminare qualcosa che non voleva condividere o che semplicemente decide in seguito di non voler più condividere. Purtroppo, non è così, in quanto su internet rimane traccia di tutto ciò che viene caricato, anche se non sembrerebbe apparentemente. Pertanto, la nascita e lo sviluppo della tecnologia rende necessaria l'esigenza da parte dello Stato di proteggere la sovranità digitale per garantire la sicurezza e la libertà delle persone online.

Proprio perché è pienamente riconosciuta l'importanza della circolazione delle informazioni nel contesto sociale in cui ci troviamo oggi, ogni individuo deve essere consapevole sia dell'inevitabilità che i propri dati circolino nel web, ma allo stesso tempo anche dei rischi che ciò potrebbe portare ai propri diritti e alla propria libertà individuale. La nascita del diritto alla protezione dei dati personali ha l'obbiettivo di bilanciare questo duplice aspetto che caratterizza la società di oggi per garantire agli individui una certa padronanza sulla circolazione delle loro informazioni. Il fatto che soggetti terzi abbiano libero accesso a questo genere di dati rende necessario fissare un quadro normativo che venga rispettato da tutti coloro che hanno a che fare con il trattamento di dati personali, al fine di evitare comportamenti illeciti. Di conseguenza, il diritto alla protezione dei dati personali si traduce nel diritto di ogni individuo di essere costantemente informato su come i propri dati vengono trattati e per quali scopi, garantendo così il controllo su di essi. Questo diritto è diventato talmente importante da essere considerato uno dei pilastri dei diritti fondamentali della personalità, i quali stanno subendo trasformazioni in linea con l'evoluzione della società.

1.2 Integrazione nell'Unione Europea delle norme sulla protezione dei dati

Inizialmente l'ottica giurisprudenziale era più spostata su un contesto economico piuttosto che dei diritti della persona, dovuto al fatto che l'Unione europea nasce come Comunità economica europea (CEE) con l'obbiettivo della libera circolazione delle persone, delle merci e dei servizi. La prima normativa che riconosce i diritti fondamentali del cittadino prende forma con il documento della Convenzione Europea dei Diritti dell'Uomo (CEDU), la quale non include

ancora la protezione della privacy come diritto fondamentale. Dopo 30 anni dalla CEDU, nel 1978 viene emanata nella Germania federale la prima legge nazionale per la protezione dei dati personali: nel contesto storico in cui l'Unione Sovietica esercitava un controllo stretto sui cittadini, si manifestava una crescente preoccupazione riguardo alla centralizzazione eccessiva delle informazioni nelle mani dello Stato, temendo che ciò potesse rafforzare la dittatura del regime. Negli anni 80, con la diffusione dei primi PC in casa viene adottata dal Consiglio d'Europa la Convenzione 108 che è il più grande documento a livello europeo per la protezione dei dati personali e che oggi si applica ai paesi terzi. Vengono evidenziati per la prima volta quattro punti importanti:

1. Definizione di dato personale come un'informazione relativa ad una persona fisica identificata o identificabile.
2. Trattamenti dei dati mediante elaborazione automatizzata.
3. Il diritto delle persone di essere informate sui trattamenti effettuati sui propri dati e sui soggetti che li hanno effettuati.
4. La qualità dei dati trattati, evidenziando l'importanza di garantire l'accuratezza e l'integrità delle informazioni.

La Convenzione 108 introduce la tutela dei dati personali come un concetto autonomo sottolineando l'importanza che, per garantire i diritti e le libertà fondamentali delle persone, il trattamento dei dati personali rispetti sempre specifici requisiti. Il 18 maggio 2018 il Consiglio ha modificato il protocollo originale, facendo nascere la Convenzione 108+, per aggiornarla e renderla più conforme ai tempi moderni, in cui le violazioni della privacy e della protezione dei dati sono diventate una questione di grande importanza.

1.3 La Direttiva 46

Negli anni 90 con la creazione di un mercato unico all'interno dell'Unione europea per la libera circolazione delle persone, delle merci e dei capitali, nasce l'esigenza di un quadro normativo a livello europeo per la protezione sulla circolazione dei dati personali. Sorge essenzialmente una sfida nel bilanciare due grandi esigenze: da un lato la libera circolazione delle persone e lo scambio di merci e capitali tra gli Stati membri, dall'altro la salvaguardia dei diritti fondamentali per i quali il concetto di privacy precedentemente noto come "diritto di essere lasciati soli" si fonde con il diritto alla protezione dei dati personali. Con l'obbiettivo di risolvere tali divergenze, la CEE adotta la Direttiva 46 del 1995 che rappresenta un vero e proprio modello europeo, e nella quale si rileva la differenza tra il diritto al rispetto della vita privata e familiare,

che si concentra principalmente sul potere di impedire intrusioni esterne, e la tutela dei dati personali che stabilisce regole sul modo in cui i dati sono trattati e dà poteri per intervenire. Si può affermare, quindi, che la Direttiva 46 del 1995 rappresenta un grande passo significativo nella protezione dei dati personali in quanto per la prima volta pone al centro il rispetto della persona e della sua privacy. La normativa, inoltre, istituisce un'autorità di controllo per garantire la corretta applicazione dei principi, delle regole e misure di sicurezza per il trattamento delle informazioni. Tuttavia, la Direttiva 46 del 1995 stabiliva i principi fondamentali che dovevano essere garantiti per la protezione dei dati personali, ma non forniva dettagli specifici sui mezzi per attuare tali garanzie. Questa mancanza di specificità ha portato a notevoli differenze nell'interpretazione e nell'applicazione della direttiva da parte degli Stati membri, compromettendo così un'efficace armonizzazione. Ad esempio, in Italia per recepire la Direttiva 46/1995 è stato necessario adottare il Codice Privacy, mentre Il Bundesdatenschutzgesetz (BDSG) è la legge federale tedesca sulla protezione dei dati in Germania. Ciò ha causato difficoltà nella gestione e nel coordinamento dei dati, sia all'interno dell'Unione Europea che a livello internazionale, obbligando le istituzioni comunitarie a lavorare ad una proposta di regolamento per rivedere la normativa sulla privacy.

1.4 Il GDPR

Nel 2016 nasce il Regolamento generale sulla protezione dei dati o GDPR ed entrato in vigore il 25 maggio 2018, che mirava a creare una disciplina sulla privacy omogenea su tutto il territorio dell'Unione europea. L'approvazione del Regolamento (UE) 2016/679 rafforza il diritto alla protezione dei dati personali ormai consolidato da tempo come un diritto fondamentale della persona, dettando le leggi riguardanti la circolazione del dato per creare un ambiente in cui i cittadini dell'UE possano sentirsi sicuri e protetti, mentre allo stesso tempo si promuove un clima di fiducia che favorisca lo sviluppo economico. Tuttavia, ancora una volta il Regolamento non entra nel dettaglio su tutti gli aspetti di applicazione della normativa, lasciando ai singoli legislatori nazionali un margine di autonomia sulle modalità di attuazione della legge.

Nel GDPR per "trattamento" si intende qualsiasi operazione o insieme di operazioni eseguite sui dati personali; si definisce "responsabile del trattamento" l'individuo, l'ente, l'amministrazione pubblica, il servizio o altra entità che stabilisce le finalità e i metodi per il trattamento di tali dati; si definisce "autorizzazione dell'interessato" qualunque espressione di volontà libera, specifica, consapevole e chiara da parte dell'interessato con la quale egli esprime il proprio consenso affinché le informazioni che lo riguardano siano sottoposti a trattamento.

Si denota una grande differenza con la precedente disciplina normativa sulla privacy, la quale era orientata essenzialmente su un approccio statico basato sul diritto della persona di escludere interferenze esterne dalla propria vita privata negando semplicemente l'accesso ai propri dati da parte di terzi. Con l'avvento del Regolamento (Ue) 2016/679 non si tratta più solo di proteggere la persona fisica, ma anche la circolazione dei suoi dati a causa dell'ampia libertà di circolazione delle informazioni attraverso i confini digitali. **L'EDPB è la massima autorità di controllo per l'applicazione delle norme dettate dal GDPR nell'UE** sul trattamento dei dati personali da parte di aziende o altre organizzazioni.

- **Doveri dei titolari del trattamento**

1. I siti web che trattano dati personali di qualsiasi genere dei loro visitatori devono esplicitamente chiedere il consenso di questi ultimi.
2. Monitorare e registrare tutte le attività di trattamento dei dati personali.
3. È importante che sia chiara la tipologia di dati trattati, lo scopo del trattamento e le eventuali trasmissioni a terze parti o paesi al di fuori dell'UE, utilizzando un linguaggio chiaro e accessibile.
4. Queste informazioni devono essere sempre consultabili dal visitatore come parte dell'informativa sulla privacy.
5. Il consenso dell'utente deve essere, in tutti i casi, libero, specifico, informato e inequivocabile; pertanto, non può essere dedotto dal semplice scorrimento di una pagina web o dalla continuazione della navigazione.
6. Valutare i possibili rischi e mitigarli adottando misure adeguate prima di procedere con il trattamento stesso.

Nei casi in cui il responsabile del trattamento non adempia a questi doveri, l'autorità di controllo può prendere delle misure correttive come ad esempio l'ammonimento, la limitazione o il divieto del trattamento. È importante tenere presente che non è sufficiente rispettare le norme in materia di protezione dei dati personali, ma i titolari del trattamento dovranno dimostrare di essere consapevoli delle modalità di trattamento dei dati e di conservazione degli stessi, insomma dovranno saper rendere conto di quanto fanno.

- **Diritti dell'interessato**

1. Portabilità dei dati, ovvero poter trasferire i propri dati da un titolare del trattamento ad un altro, compresi i social network.
2. Poter chiedere l'accesso ai dati personali che lo riguardano.

3. Oblio, vale a dire che gli individui possono decidere in qualunque momento di non voler più condividere i propri dati personali e nel momento in cui ritirano il loro consenso, il titolare del trattamento è obbligato ad interrompere immediatamente l'elaborazione delle loro informazioni ed eliminarle. È fondamentale che ciò possa avvenire con la stessa facilità con la quale l'individuo aveva dato il suo consenso iniziale.
4. Deve esserci un bilanciamento tra gli interessi del terzo e i diritti fondamentali della persona alla quale si riferiscono i dati.

1.5 Tecnologie di tracciamento dei dati

Uno dei principali strumenti di tracciamento dei consumatori sono i cookie web, ovvero piccoli file di testo che i siti web installano sul dispositivo dell'utente tramite il browser con lo scopo di tracciare le attività degli utenti attraverso le loro interazioni con il sito. Quando un cookie viene installato su un dispositivo conseguentemente all'accesso ad un sito web, la sua presenza persiste ogni volta che l'utente ritorna su quel sito, permettendo al sito di riconoscere l'utente e creare la sua storia di ricerca. Di fatti, sfruttando la cronologia delle ricerche, l'algoritmo della piattaforma è in grado di classificare gli utenti raggruppandoli in target con interessi comuni. Questo permette di creare annunci personalizzati con l'obiettivo finale di incentivare gli utenti ad effettuare un acquisto. Da qui nasce un grande punto interrogativo, ovvero se gli utenti siano maggiormente interessati alla protezione della propria privacy o all'utilizzo di uno strumento semplice ed efficace che gli fornisca in pochi secondi il prodotto che loro stanno cercando senza fare alcuno sforzo, ovvero senza perdere tempo con le ricerche nel web.

Ancor prima dell'entrata in vigore delle regolamentazioni sulla privacy, erano presenti degli strumenti basati sul browser che davano la possibilità agli utenti di limitare o cancellare i cookie relativi ad una sessione nel momento in cui abbandonavano un sito web. Di conseguenza, non appena l'utente si collegava nuovamente a quel sito venivano installati sul browser nuovi cookie che lo identificavano con un ID diverso, non permettendo alla piattaforma di poterlo identificare attraverso i dati che l'utente aveva lasciato nella sessione precedente. Tuttavia, anche se i dati non erano più riconducibili direttamente ad un particolare utente, venivano comunque inviati e memorizzati nel momento della loro generazione. L'effetto che veniva prodotto era quello di "offuscare" la storia di ricerca del singolo utente che metteva in difficoltà la piattaforma nel prevedere il comportamento all'acquisto del consumatore, in quanto ogni sessione effettuata dallo stesso utente veniva percepita dalla piattaforma come se fossero

consumatori differenti. La normativa della Direttiva sulla privacy del 2009 è stata la prima ad introdurre l'obbligo da parte delle aziende di chiedere agli utenti il consenso esplicito per l'utilizzo dei cookie o altre tecnologie di tracciamento che non sono strettamente necessarie per la fornitura di un servizio richiesto dall'utente. Successivamente, il GDPR ha introdotto il regime opt-in, un nuovo strumento con un ambito di applicazione più ampio che si applica a qualsiasi trattamento dei dati personali all'interno dell'UE. Il regime opt-in stabilisce che il consenso al trattamento dei dati deve essere esplicito, informato e libero: quando un utente visita un sito web, deve essere informato chiaramente dell'uso dei cookie o di altri strumenti di raccolta dati e deve dare il proprio consenso attivo, altrimenti la piattaforma non potrà mai procedere al loro tracciamento. Questo strumento offre agli utenti un maggiore controllo sui propri dati personali poiché, negando esplicitamente il consenso al tracciamento, impedisce la creazione di qualsiasi traccia dei loro dati su un sito. In questo modo i dati non vengono mai raccolti né memorizzati. Se i dati non vengono visualizzati affatto, previa negazione al tracciamento, non vengono confusi con i dati di coloro che non aderiscono alla politica della privacy. A differenza degli strumenti basati sulla privacy, ciò migliora la capacità della piattaforma di prevedere il comportamento all'acquisto dei consumatori perché, nonostante abbia a disposizione una quantità minore di dati, le informazioni provenienti dai consumatori che non hanno acconsentito alla protezione della privacy non vengono offuscate da coloro che aderivano alla protezione della propria privacy. Di conseguenza, per la piattaforma sarebbe più facile collegare i dati che arrivano al sito web al singolo consumatore.

1.6 Gli effetti del GDPR

Successivamente all'entrata in vigore del GDPR è stato interessante capire l'impatto che la regolamentazione ha avuto sul comportamento dei consumatori online e, di conseguenza, sui profitti delle imprese. A tale proposito, in un articolo di Guy Aridor, Yeon-Koo Che e Tobias Salz del 2020 viene eseguita un'analisi utilizzando i dati forniti da un intermediario anonimo, il quale raccoglie tutte le query di ricerca e di acquisto dei consumatori su un gran numero di agenzie di viaggio situate negli USA e nell'UE. Le query riguardanti uno specifico sito web consentono di vedere il numero di acquisti, i consumatori che li hanno effettuati e la data in cui è avvenuto l'acquisto per quello specifico sito, aggregando tutti i dati a livello settimanale. L'esperimento è basato sulla strategia nota come "difference-in-difference" (DID), che permette di misurare l'effetto del GDPR isolando l'impatto diretto della normativa. Il metodo confronta un gruppo di trattamento, costituito dai Paesi europei che hanno adottato il regolamento, con un gruppo di controllo, rappresentato dai Paesi statunitensi che non lo hanno

implementato. In questo modo, si possono isolare le differenze nell'impatto sui siti web dovute dovuto esclusivamente all'introduzione del GDPR. L'obbiettivo è capire quanto e come il GDPR abbia cambiato la gestione della privacy online, confrontando i dati delle diverse agenzie e analizzando le differenze nei risultati tra Paesi che seguono e non seguono il regolamento.

Viene stimata una regressione in cui la variabile dipendente rappresenta l'impatto che si vuole analizzare nel tempo; in particolare, la stima viene fatta settimanalmente a partire dalla data di conformità del GDPR, ovvero il 25 maggio 2018. I fattori che influenzano questo impatto sono identificati da diverse variabili, tra cui:

- c: il Paese
- j: il sito web specifico
- o: il sistema operativo utilizzato
- b: il browser web impiegato
- p: il tipo di prodotto
- t: la settimana dell'anno

In questa analisi, si cerca di comprendere come ciascuno di questi fattori possa contribuire all'impatto complessivo, tenendo conto delle loro interazioni e variazioni nel tempo e nello spazio e considerando anche la di versa tempestività di adesione alla normativa dei diversi Paesi europei.

$$y_{tcjobp} = \alpha_t + \delta_{jc} + k_c + \varepsilon_j + \gamma_o + \sigma_b + \omega_p + \beta(EU_j \times after) + \epsilon_{tcjobp}$$

Nella regressione, ***EUj*** rappresenta il sito web specifico, mentre la variabile ***after*** indica se, nella settimana corrente che si sta analizzando, il sito web ha già aderito alla normativa. Poiché i Paesi europei hanno aderito al regolamento in momenti diversi, è importante specificare, per ogni settimana, se il sito web ha già conformato alle nuove disposizioni o meno. Tutte le altre variabili, invece, vengono utilizzate per identificare **effetti fissi** associati ai vari fattori, che permettono di controllare le variazioni specifiche legate a ciascuno di essi.

1.6.1 Risposta dei consumatori

In primo luogo, si analizza il comportamento dei consumatori di fronte alla scelta di proteggere la propria privacy attraverso la possibilità di selezionare il regime opt-out, osservando quanti cookie sono stati tracciati dall'intermediario con l'obiettivo di vedere chi opta per il rifiuto del consenso e chi, al contrario, decide di concederlo. Naturalmente, è possibile tracciare solo i

consumatori che non aderiscono al regime opt-out, poiché quelli che decidono di aderire al regime risultano come 'inesistenti' per il sito web. Pertanto, per stimare il numero di utenti che continuano ad essere tracciati, si procede a calcolare la differenza tra il numero totale di consumatori e quelli che hanno adottato il regime opt-out.

$$y_t^{OBS} = U_{jt}^{TRUE} - U_{jt}^{OPT-OUT}$$

Dove $U_{jt}^{OPT-OUT}$ vale zero nel gruppo di controllo e nel gruppo di trattamento nel periodo antecedente all'implementazione del GDPR.

Risolvendo la regressione considerando come variabile dipendente il totale di cookie identificati nei siti web, si giunge al risultato che il GDPR ha causato una riduzione di circa 12,5% dei cookie totali. Questo suggerisce che un certo numero di consumatori ha scelto di non essere tracciato dalle piattaforme. Tuttavia, questa riduzione non corrisponde esattamente al 12,5% dei consumatori, poiché la variabile utilizzata per la stima è il numero totale di cookie, e non il numero di consumatori. Infatti, ogni consumatore può essere tracciato attraverso più cookie. Inoltre, utilizzando la stessa regressione, è stata stimata anche la variabile dipendente rappresentata dal numero totale di ricerche registrate dall'intermediario. I risultati mostrano una riduzione del 10,7% nel numero complessivo di ricerche registrate. Questo calo suggerisce che l'adozione del GDPR ha avuto un impatto simile anche sulle ricerche, con una diminuzione che si allinea con la riduzione dei cookie.

In seguito, sugli utenti che vengono tracciati si vuole rilevare la persistenza dell'identificatore, ovvero se lo stesso numero di cookie visti in una data settimana all'interno di uno specifico sito web ritorna dopo un certo numero di settimane all'interno dello stesso sito. Viene stimata una variabile che misura la persistenza dell'indicatore, con l'obiettivo di analizzare se lo stesso numero di cookie continua a essere tracciato in periodi successivi, espressi in settimane, fornendo così un'indicazione della durata e della continuità del tracciamento.

$$persistenza_{kt} = \frac{\left|C_{j,t} \cap C_{j,t} + k\right|}{\left|C_{j,t}\right|}$$

Dove k indica le settimane e $C_{j,t}$ l'insieme dei cookie visti nella settimana t sul sito web j. Quello che si denota attraverso il modello è un aumento repentino, subito dopo l'entrata in vigore del GDPR, della presenza dei consumatori all'interno dei siti web nei Paesi europei; di contro, non si nota nessuna differenza nei siti web statunitensi. dopo l'entrata in vigore del GDPR, ovvero una maggiore persistenza dell'identificatore online. Questo fenomeno può essere spiegato attraverso due possibili ipotesi:

1. L'ipotesi del consenso selettivo, la quale sostiene che i consumatori che rifiutano il consenso all'archiviazione dei propri dati sono quelli che visitano i siti web con minore frequenza. Di conseguenza, i consumatori che effettuano ricerche online più frequentemente e non optano per la protezione della propria privacy risulteranno maggiormente visibili.
2. L'ipotesi della sostituzione dei mezzi di tutela della privacy, la quale si riferisce al cambiamento nel modo in cui vengono gestiti i dati degli utenti che utilizzano strumenti di protezione basati sul browser. In passato, la loro cronologia di ricerca veniva semplicemente offuscata, influenzando anche i dati degli utenti che non adottavano tali strumenti. Con l'introduzione del regime opt-in del GDPR, i dati degli utenti che aderiscono alla protezione della privacy vengono invece eliminati del tutto. Questo permette agli intermediari di tracciare con maggiore precisione i dati degli utenti non protetti, creando l'impressione di una maggiore presenza di questi ultimi, anche se il loro comportamento di ricerca non è cambiato. In realtà, la maggiore visibilità deriva dal fatto che i dati non sono più "contaminati" dall'offuscamento dei dati degli utenti protetti.

È stato dimostrato che la seconda ipotesi è la più plausibile: prima dell'introduzione del GDPR, gli strumenti di tutela della privacy basati sul browser generavano un gran numero di utenti singoli con una presenza online artificialmente breve, creando un'illusione di volatilità. Con l'entrata in vigore del GDPR, il numero di questi utenti "temporanei" è drasticamente diminuito, portando ad una maggiore stabilità nel tracciamento ed ad una distribuzione più uniforme degli utenti.

1.6.2 Pubblicità online

La “previsione” consiste nell’anticipare il comportamento all’acquisto del consumatore, cioè se un consumatore acquisterà da uno specifico sito web dopo aver effettuato una data ricerca online; quindi, l’accuratezza della previsione dipende dalla storia di ricerca online del singolo consumatore. In base alle previsioni sul comportamento di acquisto dei consumatori, gli inserzionisti decidono se pubblicare o meno determinati annunci. In altre parole, la decisione di mostrare un annuncio dipende dal valore che si ritiene possa essere percepito dagli utenti, influenzando così la scelta degli inserzionisti su quali annunci presentare. In questo contesto, viene stimata una regressione per esaminare se la riduzione dei cookie e delle ricerche online, osservata nei passaggi precedenti, sia associata a una diminuzione nel numero di annunci pubblicati dagli inserzionisti. L'obiettivo è verificare se il calo dell'attività di tracciamento degli

utenti porti, di conseguenza, a una riduzione delle opportunità pubblicitarie offerte. Effettivamente, ciò che si denota è proprio un calo del numero di annunci; tuttavia, tale diminuzione non è significativa come quella dei cookie.

Per analizzare i profitti delle imprese, è necessario esaminare l'andamento del numero di click generati, in quanto rappresentano uno dei principali determinanti dei profitti delle piattaforme. Si è dimostrato che il numero di click sugli annunci online ha subito una diminuzione significativa del 13,5%, attribuibile alla riduzione dei cookie e al minor numero di ricerche. Dunque, verrebbe immediato pensare che la minor quantità di dati dovuta al GDPR causi una diminuzione delle capacità predittive dell'algoritmo della piattaforma. Tuttavia, attraverso l'analisi statistica con modelli diff-in-diff delle variabili chiave per la previsione è emerso che, complessivamente, il GDPR ha avuto un impatto positivo sulla capacità di predire il comportamento d'acquisto dei consumatori. Ciò sembra essere legato al maggiore valor medio del consumatore post-GDPR, poiché quando gli utenti aderiscono al regime opt-in del GDPR, tutte le loro informazioni vengono eliminate dal set di dati tracciati dalla piattaforma. Di conseguenza, i consumatori rimanenti risultano più facilmente rintracciabili, permettendo agli inserzionisti di collegare con maggiore precisione gli acquisti ai singoli utenti. Questo fenomeno incrementa il valor medio di ogni utente, poiché, nonostante la riduzione del volume complessivo dei dati, i dati rimanenti sono di qualità superiore. Diventa quindi più semplice identificare e targetizzare i consumatori, rendendo le campagne pubblicitarie più efficaci e contribuendo al rialzo dei ricavi delle piattaforme.

1.6.3 Traffico degli utenti online

Un'analisi dell'impatto del GDPR sul traffico web è stata condotta dalla rivista Elsevier nel 2022, basandosi sui dati di traffico provenienti dai siti web più attivi in Europa e negli Stati Uniti. L'analisi è stata fatta utilizzando come fonte principale di dati il servizio di SimilarWeb che monitora le interazioni degli utenti con i vari siti, in particolare:

1. Totale delle visite effettuate ad un dominio.
2. Durata media in secondi di ogni visita, basata sul tempo che intercorre tra la prima e l'ultima visualizzazione della pagina.
3. Numero di utenti che escono dal sito web dopo aver visualizzato la prima pagina.
4. Numero medio di pagine visualizzate per ogni visita.

Tra questi dati, le visite sono quelli di maggiore interesse per l'analisi. SimilarWeb, infatti, fornisce una distinzione dettagliata delle fonti di traffico, suddividendo gli accessi in diverse categorie. In primo luogo, vi è il traffico diretto, che si verifica quando un utente digita direttamente l'URL di un sito web nel browser. Accanto a questo, troviamo il traffico indiretto, che include visite provenienti da altre fonti, come e-mail, pubblicità, o link da siti terzi. Inoltre, è utile fare una distinzione tra ricerche organiche e ricerche a pagamento: le prime si riferiscono ai risultati che appaiono gratuitamente nei motori di ricerca come Google, senza alcun pagamento; mentre le altre riguardano gli annunci sponsorizzati che vengono visualizzati in cima o nella sezione laterale dei risultati di ricerca. Vengono selezionati come set di dati i siti web che coprono una fetta importante del traffico online nei paesi analizzati negli USA e nell'UE, sia prima che dopo il GDPR, arrivando ad un totale di 4957 siti web. A causa di oscillazioni giornaliere e mancanza di informazioni per alcuni siti e paesi, sono stati aggregati i dati settimanalmente, per un totale di 94 settimane che vanno dal 5 gennaio 2018 al 25 ottobre 2019. Viene nuovamente utilizzata la metodologia del dff-in-diff dove i gruppi di trattamento e di controllo si riferiscono rispettivamente agli utenti dell'Unione Europea, soggetti alla normativa GDPR, e agli utenti provenienti da paesi al di fuori dell'UE, ai quali il GDPR non si applica. I dati raccolti consentono di identificare la derivazione geografica degli utenti che navigano su un sito web indipendentemente dal paese di registrazione del dominio; di conseguenza, non è rilevante considerare il dominio del sito web osservato, poiché un singolo sito è obbligato a trattare in modo differenziato i suoi utenti in base alla loro provenienza geografica. Pertanto, ogni sito web specifico può essere trattato parzialmente come gruppo di trattamento o di controllo se il suo traffico deriva sia dagli USA che dall'UE.

Si stima la seguente regressione:

$$y_{i,k,c,t} = \beta_0 + \beta_1^S Post_t^S + \beta_1^L Post_t^L + \beta_2 EU_c + \beta_3^S Post_t^S \times EU_c + \beta_3^L Post_t^L \times EU_c + \alpha_i + k_k + \varepsilon_{i,k,c,t}$$

in cui la variabile $y_{i,k,c,t}$ analizza l'impatto sul traffico online del sito web "i" nel periodo precedente o successivo all'entrata in vigore del GDPR, per il paese "c", alla data della settimana "t". Nella regressione, è fondamentale distinguere tra l'impatto a breve e a lungo termine derivante dall'entrata in vigore del GDPR, poiché gli effetti della normativa potrebbero manifestarsi in modo diverso nel corso del tempo. In particolare, nel breve periodo, che va dal 25 maggio 2018 fino alla settimana precedente il 21 gennaio 2019 (quando Google ha ricevuto la multa), l'effetto del GDPR potrebbe essere meno significativo. Al contrario, nel lungo termine, l'impatto della normativa potrebbe risultare più evidente e amplificato. Per questa ragione, la variabile "Post" viene utilizzata per

catturare questa differenza temporale. Inoltre, la variabile "yS" rappresenta l'effetto casuale a breve termine, mentre "yL" riflette l'impatto nel lungo periodo.

Per prima cosa, bisogna è stata testata l'ipotesi di partenza sull'equità delle tendenze del traffico web nell'UE e negli USA prima dell'attuazione della normativa stimando la seguente equazione:

$$y_{i,k,c,t} = \beta_0 + \tau_t + \alpha_i + k_k + \varepsilon_{i,k,c,t}$$

Analizzando la variazione dei coefficienti associati alle principali variabili di interesse (traffico web, durata media della visita, numero medio di pagine visitate e tasso di abbandono del sito (utenti che lasciano il sito dopo aver visitato solo la prima pagina) si osserva che, prima dell'introduzione del GDPR, i trend di tutte queste variabili erano pressoché simili sia nell'Unione Europea che negli Stati Uniti. Dopo l'entrata in vigore del GDPR, il traffico web nell'Unione Europea ha subito una diminuzione. In particolare, nel breve periodo si è registrato un calo di circa il 4%, mentre nel lungo periodo l'impatto è stato molto più marcato, con una riduzione del traffico web pari al 15,7%. Per quanto riguarda invece il numero medio di pagine visitate e il tasso di abbandono (utenti che lasciano il sito dopo la prima pagina), le riduzioni si sono manifestate esclusivamente nel lungo termine. La diminuzione delle interazioni online da parte dei consumatori sembra indicare una crescente consapevolezza riguardo alla protezione della propria privacy. Di conseguenza, molti utenti hanno scelto di ridurre l'utilizzo di siti web percepiti come più invadenti in termini di raccolta dati. In sostanza, si può affermare che il traffico USA è un buon gruppo di confronto per valutare l'impatto del GDPR sui siti europei in quanto ha condotto l'analisi alla conclusione che il GDPR ha influenzato negativamente il coinvolgimento degli utenti online. Tuttavia, la diminuzione significativa del traffico non è stata immediata, ma si è manifestata circa un anno dopo. Si ritiene che ciò sia dovuto al tempo necessario sia per gli utenti che per le aziende per adattarsi e conformarsi alle nuove normative sulla privacy. Inoltre, si ipotizza che un fattore decisivo nel sensibilizzare le aziende sull'importanza della privacy sia stata la multa inflitta a Google il 21 gennaio 2019; infatti, dopo questa sanzione il trend del traffico nell'UE ha subito un notevole impatto negativo.

1.7 Le nuove legislazioni

Si apre, pertanto, il decennio digitale in cui le nuove tecnologie consentono ai cittadini nuove libertà e nuovi diritti che vanno al di là dell'aspetto fisico della comunità. L'obiettivo di questo nuovo decennio è creare una rete digitale dalla quale nessuno venga escluso, per aumentare la

connettività tra le persone ed agevolare le operazioni aziendali. Si vuole creare un mercato unico per i dati al quale chiunque può avere accesso dal proprio dispositivo e possa condividere informazioni su internet.

1.7.1 Data Act e Data Governance Act

Entrano in vigore due proposte legislative rispettivamente il 22 giugno 2023, il Data Governance Act, e il 22 dicembre 2023, il Data Act. La loro adozione proviene dalla consapevolezza da parte dell'UE del crescente peso dei dati digitali all'interno del contesto economico, e sono il risultato di un processo lungo e complesso nel quale c'è la necessità di proteggere i dati da coloro che li vogliono sfruttare per avere sempre più potere. Questi due regolamenti fanno parte della "Strategia Europea per i Dati" che getta le basi per la costruzione di un'economia nella quale tutti abbiano egualmente accesso ai dati digitali, dai quali trarne vantaggio. Il Data Governance Act ha l'obbiettivo di semplificare e rendere più sicura la condivisione dei dati tra enti pubblici e privati creando regole per il riutilizzo dei dati, sempre garantendo il rispetto della privacy e la protezione delle informazioni personali. Il Data Act, invece, si concentra soprattutto su come le persone e le aziende possono accedere ai dati generati da dispositivi connessi (come smartphone, elettrodomestici smart, ecc.), e introduce regole per rendere i dati più facilmente utilizzabili e condivisibili tra settori diversi e tra aziende, migliorando la collaborazione e l'efficienza. Si vuole pertanto creare uno spazio europeo comune nel quale dare luogo allo scambio dei dati in totale sicurezza, rispettando la privacy degli individui. Nonostante entrambi i regolamenti prevedono disposizioni sulla protezione dei dati personali, differiscono lievemente nel campo di applicazione. Il Data Act riguarda principalmente l'accesso e l'utilizzo dei dati pubblici, cioè quelli raccolti dalle aziende; mentre il Data Governance Act si concentra su meccanismi più specifici per la condivisione più sicura dei dati. Ogni volta che un utente utilizza un dispositivo connesso ad una rete viene generata una grande quantità di dati relativi al loro funzionamento e alle modalità d'uso, che vengono condivisi su internet. Il Data Act riguarda la loro gestione, in particolare mira a garantire che tutti gli utenti possano avere accesso a questi dati attraverso l'uso dei loro dispositivi. Da ciò si evince come le nuove legislazioni rappresentino un potente strumento per l'innovazione, poiché sono atte a promuovere lo sviluppo di nuovi prodotti e servizi con carattere innovativo, ma anche migliorare quelli già esistenti. In sostanza, i principi introdotti da queste normative offrono vantaggi significativi per una vasta gamma di soggetti: le imprese possono accedere a una mole più ampia di dati, il che favorisce la concorrenza, mentre i consumatori ottengono un controllo più diretto e trasparente sui dati dal loro dispositivo.

1.7.2 Contrasto tra Data Act e GDPR

Il Data Act costituisce il secondo passo della strategia della Commissione Europea, dopo il Data Governance Act, di creare un quadro giuridico unitario per la condivisione dei dati all'interno dell'UE. Se da un lato la rimozione di barriere facilita la trasmissione dei dati tra i diversi settori economici, dall'altro lato ciò comporta nuovi obblighi per le imprese nell'assicurarne una corretta gestione. Per tale motivo con il Data Act si vogliono introdurre dei requisiti vincolanti a tutti quei produttori di servizi che forniscono l'accesso degli utenti ai loro dati con il fine di permettere ai consumatori e alle aziende un controllo rafforzato sui propri dati, avendo ben chiaro il fine ultimo di questi, chi vi può fare accesso e in quali condizioni. A questo punto ci si è posta la domanda se la nuova normativa entrasse in conflitto con l'attuale regolamento del GDPR in materia di protezione dei dati personali. Apparentemente non viene riscontrata nessuna incoerenza tra le due normative, in quanto è apparso evidente come la concezione di dato venga trattata in maniera differente nei due regolamenti. Infatti, nel GDPR ci si riferisce strettamente a "dati personali", ovvero informazioni riconducibili ad una persona nello specifico come nome, indirizzo o e-mail; nel Data Act si trova invece un significato molto più ampio e generico di "dato", che include qualsiasi tipo di informazione digitale come numeri, immagini, suoni o video, indipendentemente dal fatto che siano legati ad una persona o meno. Tuttavia, anche se in apparenza queste due definizioni potrebbero sembrare opposte, il GDPR contiene una visione talmente ampia sulla definizione di dato personale che alcune informazioni che sembrano non riguardare direttamente una persona potrebbero comunque essere considerate dati personali. Ecco che a questo punto le due discipline potrebbero arrivare a sovrapporsi: le Autorità europee hanno manifestato nel febbraio 2022 un parere congiunto su alcune incompatibilità tra il Data Act, per quanto riguarda il corretto utilizzo e accesso ai dati, e il GDPR, in materia di protezione della privacy, evidenziando le principali criticità. La prima incompatibilità riguarda l'aspetto di portabilità dei dati, ovvero il diritto che permette alle persone di ricevere i propri dati personali in maniera leggibile da un computer e di trasferirli da un'azienda all'altra. Il divario consiste nel fatto che il Data Act prevede che i dati, personali o non personali, possano essere condivisi tra diverse aziende senza necessariamente richiedere un consenso esplicito da parte del soggetto a cui appartengono. L'obiettivo principale è favorire la libera circolazione dei dati all'interno del mercato europeo, promuovendo interoperabilità e scambi agevolati. Al contrario, il GDPR impone che il trasferimento dei dati personali da un ente all'altro avvenga solo previo consenso esplicito dell'interessato, in modo da garantire la tutela della privacy. Qualsiasi condivisione di dati senza tale consenso rischia di configurarsi come una violazione della normativa sulla protezione dei dati personali. Un'ulteriore differenza

fondamentale riguarda la definizione di "utente". Nel Data Act vengono considerati utenti sia le persone fisiche che quelle giuridiche, ossia aziende o enti. Ciò implica che sia gli individui che le imprese abbiano il diritto di accedere a qualsiasi dato generato dall'utilizzo di un prodotto, indipendentemente dal consenso del titolare dei dati. Questo approccio contrasta nettamente con il GDPR, che richiede il consenso esplicito del soggetto interessato per la condivisione e il trattamento dei dati personali. Dunque, nasce il timore che la nuova proposta di regolamento possa indurre a vedere i dati personali come un semplice bene commerciale, facilitandone l'accesso e lo scambio e violando i diritti alla riservatezza e alla protezione della privacy disciplinati dal GDPR. La soluzione proposta dalle autorità competenti prevede innanzitutto una chiara distinzione tra le situazioni in cui l'utente coincide con l'interessato, cioè la persona alla quale i dati si riferiscono, e quelle in cui l'utente non è l'interessato, come nel caso di un'azienda che accede ai dati personali di un individuo. Inoltre, si raccomanda che l'accesso ai dati sia consentito solo se conforme alle disposizioni del GDPR, garantendo così il pieno rispetto dei diritti alla privacy e alla protezione dei dati personali.

1.7.3 Digital Services Act e Digital Market Act

Il Digital Services Act (DSA) è un'importante regolamentazione entrata in vigore il 23 agosto 2023 che introduce norme più rigorose sulla circolazione dei contenuti delle piattaforme digitali, sulla loro trasparenza e la protezione dei diritti degli utenti. Tale regolamento viene applicato insieme al regolamento gemello del Digital Market Act (DMA), il quale impone una serie di obblighi e divieti su quelle piattaforme digitali identificate come "gatekeeper" per garantire equità e concorrenza nei mercati digitali dell'UE. Nel contesto del DMA, per gatekeeper ci si riferisce a quelle piattaforme digitali che costituiscono una dimensione significativa nel mercato, con una grande base di utenti e una posizione consolidata. Le due legislazioni formano insieme la legge sui servizi digitali, la quale essendo ampiamente applicabile comprende una vasta gamma di fornitori di servizi che trasmettono o memorizzano informazioni, comprese le piattaforme di social media. Le norme del DSA vengono applicate gradualmente con ordini di livelli differenti che dipendono dalla dimensione del servizio: per esempio, le infrastrutture di rete e i servizi di hosting si vedono applicati degli obblighi di livello inferiore rispetto invece alle piattaforme digitali alle quali vengono imposte le norme più severe.

L'obiettivo è individuare ed eliminare i contenuti illegali attraverso l'uso di strumenti noti come "segnalatori attendibili". Questi strumenti sono impiegati per garantire la sicurezza e la trasparenza dei contenuti sulle piattaforme digitali, contribuendo a mantenere un ambiente online più sicuro e affidabile. Tale funzione viene svolta spesso da organizzazioni o esperti che

hanno il compito di identificare e segnalare la presenza di contenuti illegali che violano le politiche delle piattaforme, come incitamenti all'odio, abuso sui minori, terrorismo, manipolazione o distorsione delle informazioni. La necessità di regolamentare è emersa con la crescita dei giganti tecnologici, che hanno trasformato i modelli tradizionali di creazione del valore attraverso l'innovazione basata sui dati. Le piattaforme digitali si concentrano principalmente sull'estrazione di dati da ogni aspetto della vita individuale e sociale degli utenti, facendo del loro business la trasformazione e la rivendita di questi dati sotto forma di informazioni. Questo processo di trasformazione conferisce ai dati un nuovo valore, rendendoli una risorsa centrale nell'economia digitale. Fino a poco tempo prima le grandi piattaforme si autoregolamentavano, determinando autonomamente il ruolo e le opportunità sia degli utenti individuali che delle aziende. Tuttavia, questo approccio ha portato a una situazione in cui, a causa delle problematiche sopra descritte, è diventato urgente imporre norme più severe sui contenuti e sui comportamenti delle Big Tech, al fine di tutelare i diritti degli utenti e garantire un ambiente digitale più equo e sicuro.

Sebbene il traffico online sia diminuito significativamente, comportando per le aziende una riduzione delle visite ai siti web e maggiori costi, poiché hanno meno opportunità di sfruttare i dati degli utenti a fini economici, è stato dimostrato che il guadagno complessivo in termini di privacy per gli utenti è nettamente superiore e compensa ampiamente la perdita di profitto delle imprese. Questo perché la riduzione del traffico è il risultato di una maggiore consapevolezza e informazione da parte degli utenti riguardo ai termini della loro privacy, che li porta a evitare consapevolmente i siti in cui la protezione dei loro dati non è garantita.

1.8 Contrasto tra economisti e studiosi della privacy

Con il passare del tempo, sono emerse diverse concezioni della privacy, che hanno portato a differenti modalità di percepirne l'efficacia e il valore nelle regolamentazioni. Il dibattito riguardo il ruolo della privacy nell'economia è iniziato a sorgere verso la fine degli anni 70' e l'inizio degli anni 80', quando si scontrarono principalmente due visioni: quella degli economisti e quella degli studiosi della privacy. Le varie sfaccettature vengono analizzate in profondità in un documento di Alessandro Acquisti dell'Università Carnegie Mellon: l'autore traccia un percorso che parte dalla concezione tradizionale della privacy, specialmente dal punto di vista economico, esplorandone le trasformazioni nel contesto del mercato digitale e della crescente competitività economica. Nel mondo moderno la privacy ha acquisito lo status di diritto fondamentale strettamente legato a valori come la libertà e l'autonomia e, con il rapido sviluppo dell'economia dei dati, non è più soltanto un diritto da garantire, ma viene percepita come un

valore da proteggere attivamente. Di conseguenza, si è cominciato a parlare di "gestione dei rischi" per la privacy, riflettendo le nuove sfide emerse dall'espansione dell'economia digitale e dall'uso massiccio dei dati personali.

Tuttavia, gli studi economici hanno sempre avuto una visione molto ristretta di privacy, pensandola essenzialmente come un insieme di scelte individuali, e si basano sull'analizzare le inefficienze economiche e il maggior numero di costi che questa reca alla società, escludendo dalla loro analisi, invece, i potenziali danni che una mancanza di privacy porterebbe ai consumatori. L'economista Richard Posner, esponente della scuola di Chicago, ha descritto la privacy come un'invenzione moderna concepita per celare informazioni che potrebbero essere utili negli scambi di mercato, attribuendole la responsabilità di inefficienze economiche. Secondo Posner e altri economisti della sua scuola di pensiero, come George Stigler, i consumatori prendono decisioni sulla privacy in modo razionale, guidati da interessi economici: essi condividono volentieri informazioni che li avvantaggiano, mentre tendono a nascondere quelle sfavorevoli. Da questa prospettiva, le scelte di privacy appaiono orientate e stabili, con la condivisione di dati personali che riflette le preferenze dei consumatori. Inoltre, Posner e Stigler ritenevano che la privacy fosse un ostacolo alla raccolta e all'utilizzo efficiente delle informazioni, in quanto i consumatori, a loro parere, condividono solo i dati necessari per ottenere vantaggi economici.

Gli economisti si concentrano soprattutto sulla protezione dei dati personali poiché è un aspetto più facilmente quantificabile, trascurando la dimensione più astratta e psicologica del controllo sui propri dati, che va oltre motivazioni economiche. Questo limite interpretativo li porta a considerare la privacy esclusivamente come uno strumento per ottimizzare le preferenze dei consumatori e facilitare la discriminazione dei prezzi ignorando altre implicazioni, come i rischi legati alla sicurezza dei dati. Nella visione economica dominante, la privacy assume il ruolo di un bene intermedio, ovvero un mezzo che le persone gestiscono per ottenere vantaggi economici. Tuttavia, alcuni economisti ammettono che essa possa essere anche un bene finale, apprezzato per ragioni personali o di gusto, indipendentemente dai benefici materiali che ne derivano. Nel 1980 Jack Hirshleifer ha ampliato il dibattito andando oltre gli aspetti puramente quantitativi sostenuti dagli economisti tradizionalisti, evidenziando come la privacy fosse strettamente connessa a valori quali la libertà, la dignità e l'autonomia personale. A partire dagli anni '90, una nuova generazione di economisti, tra cui Hal Varian, Eli Noam e Kenneth Laudon, ha adottato una prospettiva più ampia, considerando la privacy non solo un tema economico, ma anche un diritto fondamentale legato alla libertà e all'autonomia individuale nella società. Questa visione riconosce la privacy come un processo dinamico di gestione dei

confini personali, che implica decidere cosa condividere con il mondo esterno e cosa proteggere. Le due principali prospettive sulla privacy si concentrano rispettivamente sull'occultamento e sul controllo: mentre l'occultamento si riferisce al nascondere informazioni specifiche, il controllo riguarda l'autodeterminazione e la capacità di gestire consapevolmente i propri dati.

Alcuni studi sul comportamento dei consumatori in relazione alla privacy dimostrano che le loro decisioni non sono affatto dettate da una logica puramente razionale in cambio di piccole ricompense; sono, invece, molti altri fattori, spesso legati alla vita quotidiana, ad influenzare le scelte delle persone riguardo alla protezione dei propri dati. In particolare, la presenza o meno di privacy induce le persone ad avere atteggiamenti differenti in base alle situazioni che potrebbero rilevarsi danneggianti per le persone stesse. In tale contesto, uno studio condotto nel 2022 dagli economisti Derksen, McGahan e Pongeluppe, nel contesto sanitario, ha analizzato la gestione dei dati dei pazienti affetti da AIDS memorizzati in cartelle cliniche elettroniche. L'obiettivo era indagare le preoccupazioni dei pazienti riguardo alla potenziale diffusione o utilizzo improprio dei loro dati sanitari per finalità poco trasparenti. In questo contesto, la privacy viene interpretata come l'opposizione dei pazienti nel consentire alla clinica l'accesso alle proprie informazioni mediche, temendo che queste possano essere usate in modo ambiguo o non etico. Lo studio ha evidenziato un impatto negativo sul benessere dei pazienti: la paura di una compromissione della propria riservatezza potrebbe dissuaderli dal sottoporsi a controlli medici, con conseguenti rischi per la loro salute.

I danni derivanti dalla violazione della privacy sono così numerosi e diversificati che risulta complesso quantificarli con precisione. Ovviamente, quelli più facili da misurare sono i danni economici, ma altri tipi di danni come quelli fisici, alla reputazione, psicologici e relazionali sono molto più difficili da valutare in termini quantitativi. Pertanto, ogni categoria presenta diverse cause e richiede approcci specifici per essere gestita correttamente. Ogni giorno le persone cercano di tutelare la propria privacy sia online che offline, adottando comportamenti che variano a seconda del contesto in cui si trovano. Questo dimostra che, contrariamente all'idea che la privacy sia una "creazione" moderna legata all'avvento di Internet, in realtà è una necessità intrinseca nell'essere umano. Diverse ricerche evidenziano come il bisogno di tutelare i propri spazi personali sia sempre esistito, evolvendosi insieme ai cambiamenti della società e ai nuovi bisogni che ne derivano. La protezione dei propri confini in base alle persone con cui si interagisce o all'ambiente in cui ci si trova, è un istinto naturale che si manifesta attraverso azioni semplici e spontanee, sia nel mondo digitale che nella vita quotidiana.

A causa di queste visioni opposte, anche le opinioni sulla regolamentazione della privacy risultano divergenti. Nelle scienze sociali la protezione dei dati personali è considerata fondamentale a causa del crescente valore dei dati generato dalle piattaforme digitali, e si ritiene importante introdurre regolamentazioni a tutela della privacy, viste come strumenti utili per interventi politici efficaci. Gli economisti tradizionali, al contrario, sono scettici riguardo l'efficacia delle normative in campo di privacy poiché le considerano troppo rigide e potenzialmente limitanti per i consumatori, i quali potrebbero essere ostacolati nello scambio dei propri dati in cambio di benefici economici. Dunque, gli economisti tendono a concentrarsi principalmente su danni economici misurabili, tralasciando strumenti adeguati a valutare correttamente l'impatto su altri valori. Questi valori, pur essendo essenziali per il funzionamento della società, non sono facilmente quantificabili dal punto di vista economico e, di conseguenza, le decisioni politiche rischiano di trascurare aspetti cruciali della privacy, privilegiando metriche economiche più semplici da analizzare.

1.9 Asimmetrie informative

La privacy, pur essendo considerata un bene economico, si distingue nettamente dai beni fisici: quando lasciamo i nostri dati online, è difficile prevedere come verranno utilizzati, poiché possono essere condivisi e sfruttati da diverse piattaforme senza il nostro consenso. Questa imprevedibilità genera un'asimmetria informativa che avvantaggia le aziende, rendendo complicato per i consumatori stimare il reale valore dei propri dati e prendere decisioni ottimali per limitare le esternalità negative. Inoltre, i rischi associati all'uso dei dati sono amplificati dal contesto in cui vengono impiegati. Le informazioni personali, spesso ricavate senza consenso esplicito, riguardano aspetti sensibili come salute, comportamenti e opinioni politiche, e la capacità delle piattaforme di estrarre e diffondere rapidamente questi dati, anche a terzi, solleva serie preoccupazioni per la sicurezza e la tutela della privacy, evidenziando uno squilibrio tra i profitti delle imprese e i diritti degli utenti.

Sebbene gli economisti riconoscano la presenza di asimmetrie informative, credono che bastino interventi semplici, come fornire più informazioni, educare i consumatori per risolvere queste difficoltà o chiarire come gestire al meglio la privacy online. Al contrario, la ricerca comportamentale ritiene che i problemi legati alla privacy siano molto più profondi e complessi e che non si tratta solo di educare i consumatori: le barriere informative, comportamentali ed economiche sono sistemiche e pervasive, rendendo inefficaci semplici "istruzioni" o interventi educativi. Per questo, la ricerca comportamentale sostiene che tali ostacoli strutturali non possono essere risolti con soluzioni superficiali e richiedono un approccio più completo e

incisivo per proteggere realmente la privacy degli individui. Questo è evidente nella nostra esperienza quotidiana online: ogni volta che apriamo una pagina web, siamo sommersi da lunghe e complesse informative sulla privacy, difficili da comprendere e da elaborare pienamente. Le aziende e le piattaforme sfruttano questa situazione sapendo che la maggior parte degli utenti, di fronte alla complessità e al numero di opzioni, finirà per accettare le condizioni senza rifletterci troppo. Rifiutare il tracciamento dei dati, infatti, appare spesso complicato o poco pratico, inducendo l'utente a cedere senza avere pieno controllo sulle proprie scelte. Proteggere la privacy online è molto più complesso rispetto alla vita quotidiana, dove prendiamo decisioni in modo quasi automatico, senza pensarci troppo. Al contrario, le piattaforme digitali mettono l'utente di fronte a scelte di privacy che sembrano complicate, scaricando su di lui la responsabilità di gestire un problema che non ha creato. Queste piattaforme presentano la questione come se fosse una decisione che spetta all'utente, ma in realtà lo manipolano spingendolo a fare scelte che spesso non tutelano adeguatamente la sua privacy.

In tale contesto nasce la concezione del "paradosso della privacy", dove per paradosso si intende il comportamento contraddittorio che gli utenti assumono online. In poche parole, se da un lato gli utenti si mostrano preoccupati verso il contesto privacy, dall'altro apparirebbero come se non gli importasse pur di conseguire i loro obbiettivi online, accettando regole e condizioni imposte dalle piattaforme senza preoccuparsene troppo. Ci sono opinioni contrastanti riguardo al fatto che il paradosso della privacy può davvero essere e considerato tale. Alcuni studi suggeriscono che non è necessario che ci sia una corrispondenza perfetta tra le opinioni o le attitudini delle persone sulla privacy e i loro comportamenti concreti. La relazione tra come le persone pensano alla privacy e come si comportano è troppo complessa per poter affermare che i consumatori siano incoerenti tra ciò che pensano e le azioni che compiono online. Sul paradosso della privacy ci sono opinioni divergenti: da un lato, gli studi economici sostengono che non si tratta di una questione così rilevante da richiedere interventi regolatori, minimizzando l'importanza del problema; dall'altro lato, ci sono coloro che attribuiscono il divario, definito come paradosso, tra le attitudini delle persone sulla privacy e i loro comportamenti a diversi ostacoli, come la mancanza di informazioni, difficoltà comportamentali e limitazioni economiche. Oltre a questi, tra l'altro, ci sono anche gli ostacoli economici: proteggere la privacy online può diventare talmente costoso per gli utenti che le opzioni disponibili risultano quasi irraggiungibili. Tutti questi fattori possono impedire ai consumatori di prendere decisioni informate e coerenti riguardo alla protezione della loro sfera privata.

CAPITOLO 2 - ECONOMIA DEI DATI

2.1 Big data

Negli ultimi anni i mercati digitali stanno sviluppando molto rapidamente dei modelli di business incentrati sulla raccolta e utilizzo dei dati. Connettendoci alla rete Internet abbiamo la possibilità di accedere a un'ampia gamma di informazioni, ma lasciando traccia dei nostri dati digitali. A questa mole di dati è stato dato il nome di "Big Data", ovvero una varietà talmente estesa di informazioni, personali e non personali, che per gestirli sono necessari degli strumenti più avanzati che si stanno sviluppando al giorno d'oggi e che analizzano i dati provenienti da varie fonti per trasformarli in informazioni di grande valore. L'entità di questi dati è sempre più in aumento grazie allo sviluppo di piattaforme digitali, come i social media, che giocano un ruolo fondamentale nella formazione delle nostre identità online e attraverso le quali gli utenti interagiscono con gli altri mostrando diversi tipi di contenuti. Lo scopo principale dell'accumulo dell'enorme mole di dati è quello di estrapolare da essi della conoscenza utile attraverso l'analisi delle relazioni tra le informazioni disponibili, applicabili in diversi campi come ad esempio nel marketing, nel management, nella pubblica amministrazione ed infine anche a livello scientifico. Grazie a due grandi sviluppi tecnologici, il cloud computing e il data mining, le piattaforme digitali hanno trasformato il loro modo di utilizzare i dati. Il cloud computing ha rivoluzionato il modo in cui le risorse informatiche vengono distribuite e utilizzate, permettendo una maggiore flessibilità, riduzione dei costi e di velocità nell'elaborazione dei dati. Dunque, le piattaforme possono sfruttare al meglio il potenziale di questi ultimi per generare valore economico migliorando l'efficienza operativa, ottimizzando le strategie di marketing, personalizzando l'esperienza del cliente e identificando nuove opportunità di mercato. La riduzione dei costi delle risorse informatiche genera economie di scala ed esternalità che hanno impatti significativi sull'economia, specialmente per gli utenti. Oggi il costo per l'utilizzo delle piattaforme informatiche è praticamente nullo, un esempio tangibile è rappresentato dai servizi dei social media e motori di ricerca che offrono accesso gratuito agli utenti, finanziandosi attraverso pubblicità mirate e raccolta di dati. Questo significa che gli utenti possono accedere a una vasta gamma di servizi senza dover pagare direttamente per il loro utilizzo. La riduzione del costo per gli utenti è stata resa possibile in parte grazie alla scalabilità e all'efficienza del cloud computing, che consente alle aziende di offrire servizi su larga scala a costi relativamente bassi. Da ciò che è stato detto emerge che i Big Data rappresentano un asset strategico fondamentale, soprattutto per le aziende che traggono significativi benefici economici dall'analisi e dalla commercializzazione di queste immense quantità di informazioni. In altre parole, i Big Data sono un autentico oggetto di scambio sul mercato, svolgendo un ruolo fondamentale nell'economia contemporanea poiché consentono di plasmare le caratteristiche dei prodotti offerti. Ciò è particolarmente evidente nei settori in cui l'obiettivo principale è la creazione di servizi che

consentano la profilazione individuale degli utenti, dando vita a un'offerta personalizzata su misura per ciascun consumatore. Questa capacità di profilazione consente alle aziende di adattare in modo preciso e mirato i propri prodotti e servizi alle esigenze specifiche di ogni singolo individuo, garantendo un'esperienza personalizzata e altamente rilevante per il consumatore.

La raccolta e l'accesso ai dati non rappresentano più una barriera insormontabile, grazie alla vasta disponibilità di fonti. I dati, dunque, non sono una risorsa scarsa, bensì abbondante, e ciò rende l'ingresso nel mercato del commercio dei dati relativamente accessibile anche per le nuove imprese del settore. In apparenza, chiunque può ottenere e utilizzare i dati come componente essenziale della propria catena di valore. Di conseguenza, non è la fase iniziale di raccolta a costituire il vero elemento distintivo per un'azienda, ma la capacità di estrarre dai dati informazioni significative e utili per il mercato, che conferiscono un reale vantaggio competitivo. In altre parole, il meccanismo che le grandi piattaforme, come Google e Facebook, sfruttano è dovuto agli effetti di rete, grazie ai quali le ricerche passate che vengono memorizzate nello storico della piattaforma consentono di fornire agli utenti ricerche attuali sempre più accurate. Tali condizioni consolidano significativamente la posizione di potere delle grandi piattaforme esistenti, rendendo l'ingresso di nuove aziende molto più difficile. Grazie ai dati rilasciati dagli utenti riguardo le loro esigenze specifiche, le aziende si servono di pubblicità mirate per catturare i loro interessi e promuovere i loro prodotti. Pertanto, nei mercati digitali i dati relativi agli utenti costituiscono una risorsa chiave per instaurare rapporti di fiducia solidi con la propria clientela, scoraggiandola dal provare nuovi prodotti o servizi di aziende concorrenti.

Ottenere l'accesso a questi dati con il consenso esplicito dell'utente non è l'unico aspetto rilevante; è fondamentale anche valutare per quanto tempo le aziende possano conservarli una volta raccolti. Questo implica un'attenzione specifica alla gestione dello storico dei dati, che si accumula in base alle ricerche e alle interazioni degli utenti nel tempo. Per comprendere quantitativamente i reali vantaggi che un'ampia mole di dati può conferire ad un'azienda, gli studiosi Lesley Chiou e Caterina Tucker hanno condotto un esperimento naturale che analizza statisticamente il traffico web per valutare l'effetto delle modifiche delle politiche di conservazione dei dati sui comportamenti degli utenti. In questo studio, dunque, si è considerata la variazione normativa europea sui tempi di conservazione dei dati utente come uno shock endogeno: una variazione esterna che influenza il tempo concesso alle aziende per memorizzare e utilizzare i dati, passando da 18 a 6 mesi per Bing e da 13 a 3 mesi per Yahoo!. Lo scopo dell'analisi è valutare se, e in che modo, i benefici delle imprese in termini di accuratezza delle ricerche siano influenzati dalla ridotta disponibilità di dati, confrontando i risultati ottenuti prima e dopo l'adozione delle nuove normative. L'analisi fatta si basa sui dati forniti dal sistema Experian Hitwise, il quale raccoglie informazioni sui comportamenti

online combinando i registri dei siti web forniti da provider di servizi Internet con dati provenienti da gruppi di utenti che hanno acconsentito a partecipare alla raccolta. Il campione di traffico web analizzato è costituito dai motori di ricerca Bing e Yahoo!, che rappresentano il gruppo di trattamento. Per questi motori si esamina il traffico verso altri siti web, detti "downstream", prima e dopo le modifiche alle rispettive politiche di conservazione dei dati. I risultati ottenuti vengono poi confrontati con il comportamento di altri motori di ricerca, come Google, i quali fungono da gruppo di controllo. Nell'analisi, i motori di ricerca svolgono il ruolo di "sorgenti" del traffico web: gli utenti vi accedono per effettuare ricerche e poi, seguendo i link nei risultati, si dirigono verso i siti downstream. L'obiettivo è quindi misurare l'effetto delle modifiche normative sul traffico che Bing e Yahoo! indirizzano verso i siti downstream, confrontandolo con quello generato da Google.

Tutto ciò è stato dimostrato empiricamente stimando una regressione, rispettivamente per entrambi i motori di ricerca, dove si mette in evidenza la differenza del rispettivo traffico in uscita tra il periodo antecedente e successivo ai cambiamenti nelle politiche della privacy. Ad esempio, per Bing viene stimata la % di visite utenti che, dopo aver digitato una specifica query nel motore di ricerca "j", vengono indirizzate ad un sito downstream "i", nella settimana "t":

$$\%visite_{i,j,t} = \beta_0 + \beta_1 Post_t \times Bing_j + \delta_y + \alpha_i + week_t + \varepsilon_{i,j,t}$$

Dove:

- Delta è un effetto fisso per il motore di ricerca j.
- Post identifica le settimane successive alla modifica della normativa.
- I controlli alfa sono gli effetti fissi del sito web a valle.
- Week indica gli effetti fissi settimanali che catturano la variazione sul volume delle ricerche in una specifica settimana t.
- Il coefficiente beta1 sull'interazione misura l'effetto della modifica della politica sul traffico web indirizzato da Bing.

Dopo aver eseguito la stessa analisi per il motore di ricerca Yahoo!, ciò che ne viene fuori è che l'effetto stimato di variazione del traffico è statisticamente insignificante. In altre parole, ciò significa che la modifica nella politica di archiviazione non ha avuto un effetto sulle visite a valle ai siti di ricerca. Tuttavia, è importante riconoscere i limiti di questa ricerca; innanzitutto, il campione esaminato è ristretto a Bing e Yahoo!, due motori di ricerca che non sono leader di mercato, il che potrebbe influenzare la generalizzazione dei risultati. Inoltre, lo studio si basa su un periodo di osservazione limitato, il che lascia aperta la possibilità che effetti significativi possano emergere solo nel lungo periodo.

2.2 Effetti di rete

L'effetto di rete avviene quando il valore di un determinato prodotto o servizio cresce all'aumentare del numero di utenti che lo utilizzano, viene denominato anche come economia di scala dal lato della domanda. In sostanza, l'attrattività di una piattaforma aumenta proporzionalmente al numero di utenti attivi che vi partecipano, poiché diventa più ricca di contenuti, interazioni e opportunità di connessione. Pertanto, l'effetto di rete è amplificato dagli stessi utenti della piattaforma che invitano attivamente i loro contatti a partecipare. A questo punto, la piattaforma diventa così attraente e funzionale che l'invito da parte degli utenti esistenti diventa un potente driver per la crescita esponenziale della base utenti. Questi appena descritti sono gli effetti di rete diretti, ma ci sono anche gli effetti di rete indiretti e si manifestano quando la diffusione di un prodotto dipende dallo sviluppo di un altro prodotto ausiliare. Questo secondo prodotto contribuisce ad incrementare il valore del prodotto originale, amplificando così la sua diffusione. All'inizio della divulgazione di una nuova tecnologia, inizialmente sono pochi gli individui che ne fanno reale utilizzo, e soltanto in seguito, man mano che la tecnologia si sviluppa, il numero di utenti attivi aumenta progressivamente. Nel corso degli anni, è emerso che le curve di adozione di nuovi prodotti o servizi seguono tipicamente un andamento a forma di "S". Questo perché all'inizio si manifesta un'incertezza riguardo al potenziale della nuova tecnologia, ma con il tempo, una volta superata la fase iniziale, l'adozione cresce rapidamente fino a raggiungere un punto di saturazione. Secondo questa legge, il valore di un'impresa è direttamente proporzionale al quadrato del numero di utenti che utilizzano una determinata piattaforma. In altre parole, in mercati caratterizzati dagli effetti di rete, l'aggiunta di un solo utente genera un aumento della domanda per quel servizio in modo più che proporzionale, in contrasto con i mercati tradizionali. Questo fenomeno noto come "positive feedback" favorisce le economie di scala, quindi il raggiungimento di un elevato numero di utenti che equivale a maggiori profitti per le imprese. Le nuove imprese che hanno interesse ad entrare in questo mercato dovranno affrontare elevati costi fissi d'investimento, che potranno recuperare solamente attraverso i profitti futuri; dopo aver affrontato l'investimento iniziale per entrare nel mercato e consolidare la propria posizione, il settore presenta i cosiddetti "rendimenti di scala crescenti". Queste aziende saranno inevitabilmente svantaggiate rispetto a quelle già consolidate nel mercato, le quali godono dei vantaggi delle economie di scala e conseguono profitti considerevoli. Ad esempio, piattaforme come Facebook vedono diminuire i propri costi in maniera direttamente proporzionale all'aumentare degli utenti sulla propria piattaforma. Questa è una delle ragioni per cui gli utenti sembrano usufruire del servizio offerto da Facebook in modo apparentemente gratuito.

L'effetto lock-in è una delle tante conseguenze degli effetti di rete, ovvero quando un individuo si vede costretto a continuare ad usufruire dello stesso prodotto o servizio a causa di investimenti precedentemente effettuati, anche in presenza di alternative più efficienti disponibili. Nel contesto delle piattaforme digitali, specialmente sui social media, l'utente può sentirsi intrappolato nell'utilizzo di una piattaforma specifica a causa dei legami e delle relazioni sviluppatesi nel tempo. Questo senso di intrappolamento è il risultato degli investimenti, sia monetari che emotivi, che l'utente ha fatto sulla piattaforma esistente e passare ad un servizio alternativo significherebbe perdere questi investimenti. Ciò di cui stiamo parlando sono gli switching costs, ovvero l'insieme di barriere che un utente deve superare quando decide di passare da un servizio o prodotto ad un altro. I costi da affrontare potrebbero essere:

- Finanziari: la tariffa monetaria per passare ad un nuovo servizio.
- Relazionali: perdita di relazioni con altri utenti istauratisi all'interno della piattaforma.
- Temporali: il tempo e lo sforzo necessario per adattarsi e imparare ad utilizzare un nuovo servizio o prodotto.
- Psicologici: eventuali timori riguardanti la sicurezza o la qualità del nuovo servizio.
- Rischio di perdita di dati o di contenuti personalizzati quando si cambia da una soluzione ad un'altra, che potrebbe comportare perdita di informazioni critiche.

  Insieme, questi costi creano una sorta di "prigione dietro la scelta", dove gli utenti possono sentirsi intrappolati nell'utilizzo di una particolare piattaforma per evitare di perdere tutto il valore creato o percepito. Questo fenomeno riduce la libertà di scelta degli utenti e può portare ad una maggiore fedeltà alla piattaforma attuale, anche se altre opzioni potrebbero offrire vantaggi superiori. Le grandi piattaforme dominanti possono sfruttare queste situazioni per consolidare ulteriormente il loro potere, ostacolando la concorrenza e l'innovazione, poiché limitano lo spazio per nuovi attori e idee nel mercato.

2.3 Piattaforme di transazione

Gli effetti di rete precedentemente visti possono agire dallo stesso lato del mercato o in due versanti opposti. Infatti, la crescente globalizzazione e i progressi tecnologici hanno favorito l'emergere delle piattaforme digitali multi-versanti: aziende che operano simultaneamente su più fronti interconnessi, sfruttando una delle risorse più influenti del mondo moderno, ovvero i dati degli utenti. L'influenza di questi ultimi ha effetti significativi nel tempo e nello spazio in cui operano le imprese del web. Un mercato a due parti è costituito da due gruppi distinti di

utenti che utilizzano una piattaforma comune per comunicare tra loro. Le piattaforme digitali che svolgono questo ruolo sono diverse:

- social network come Facebook e Instagram che richiedono ai propri utenti le informazioni personali per accedere e poter interagire con altri utenti;
- motori di ricerca come Google che rispondono alle richieste formulate dagli utenti nella barra di ricerca, fornendo diversi risultati e memorizzando le loro query;
- siti di e-commerce come Amazon che mostrano sul portale una vasta gamma di prodotti da offrire ai propri clienti.

Indipendentemente dal tipo di piattaforma, il loro ruolo è quello di agire come intermediario, facilitando il collegamento tra acquirenti e venditori in modo efficiente. Questi ultimi sono gli inserzionisti, ovvero soggetti che sfruttano la piattaforma intermediaria per promuovere il proprio marchio e farsi conoscere dal gruppo di utenti appartenenti all'altro lato del mercato, cioè i compratori. Il compenso che la piattaforma guadagna dagli inserzionisti per concedere loro un canale mirato per raggiungere il proprio pubblico di riferimento è direttamente proporzionale al numero di utenti che riesce a coinvolgere. In altre parole, più la piattaforma raggiunge un vasto pubblico, più gli inserzionisti sono disposti ad investire per promuovere i loro prodotti e servizi su di essa. Per quanto riguarda il lato dei compratori, apparentemente le piattaforme offrono loro servizi gratuiti per connettersi e cercare informazioni, ma in realtà il vero motore di questi servizi sono i dati personali degli utenti: le piattaforme raccolgono e registrano queste informazioni appositamente, poiché ciò consente loro di creare valore economico. Il loro ruolo è pertanto quello di intermediario per permettere il collegamento tra acquirenti e venditori in maniera efficiente. Entrambi i gruppi di utenti possono in questo modo dialogare fra di loro, producendo un certo valore per entrambe le parti e abbattendo gli elevati costi di transazione. In questo contesto, le esternalità di rete sono molto rilevanti, in quanto entrambi i versanti generano degli effetti l'uno sull'altro che vengono chiamati "cross-side network effects". Questi effetti possono avere sia una natura positiva che negativa:

- Si hanno effetti positivi quando gli utenti di un gruppo traggono beneficio dall'aumento del numero di utenti dell'altro gruppo.
- Gli effetti negativi sono dovuti all'eccessiva presenza di acquirenti che fanno pubblicità dei loro prodotti, che in alcuni casi possono infastidire gli utenti della piattaforma incentivandoli a ridurne l'uso.

Oltre ai cross-side network effects, gli utenti di un gruppo possono generare effetti anche nei confronti degli utenti del gruppo stesso, questi sono chiamati "same-side network effects" e seguono la stessa logica dei cross-side network effects.

In conclusione, il modello di business delle piattaforme multi-versanti si fonda sulle esternalità create tra i vari gruppi di utenti che la piattaforma mette in contatto e che le permettono di godere di rendimenti di scala crescenti al crescere del numero di utenti.

Come in ogni mercato, anche nelle piattaforme multi-versanti si cercano strategie per massimizzare il profitto complessivo, intervenendo su entrambi i lati del mercato. In questo contesto, si è individuato il lato più sensibile al prezzo, adottando parallelamente politiche di pricing più aggressive sull'altro lato, dove la domanda cresceva in modo più rapido grazie agli effetti di rete generati dall'interazione con il primo lato. Questa strategia consente alla piattaforma di massimizzare il profitto complessivo sfruttando le differenze nei comportamenti di pricing tra i due lati del mercato e capitalizzando sulle dinamiche degli effetti di rete. Nel caso specifico di Facebook, l'azienda ha compreso che il vero valore della piattaforma risiede nell'attrarre e coinvolgere gli utenti che visualizzano la pubblicità dei prodotti, cioè i consumatori, anziché i venditori che desiderano promuovere i loro servizi. Di conseguenza, la strategia di pricing di Facebook mira a sollecitare i consumatori ad iscriversi alla piattaforma offrendo loro un servizio senza costi monetari aggiuntivi. L'azienda sa che più utenti si iscrivono e interagiscono sulla piattaforma, maggiore è il valore per gli inserzionisti che desiderano promuovere i propri prodotti o servizi. In questo modo può addebitare tariffe pubblicitarie più elevate grazie alla vasta portata e all'attenzione degli utenti.

2.3.1 Valore generato per gli stakeholders

Sebbene la pubblicità online sembri vantaggiosa, gli effetti sono talmente complessi che richiedono un'analisi approfondita dei benefici e dei costi per ogni attore coinvolto. Uno studio della letteratura economica sulla privacy negli anni 80' ha analizzato e confrontato due scenari opposti, ciascuno con i propri vantaggi e svantaggi. Da un lato, l'idea di una maggiore privacy che si traduce in misure di regolamentazioni più severe e che può ridurre i benefici economici associati all'uso dei dati. Dall'altro lato, un minor livello di privacy che implica una regolamentazione meno rigida e una maggiore circolazione dei dati personali, che favorisce vantaggi economici concreti. L'obiettivo era capire se i benefici delle piattaforme digitali, come la riduzione dei costi di ricerca per gli utenti e l'aumento dei profitti per i commercianti, superano i costi in termini di benessere. La pubblicità comportamentale favorisce tutti gli attori

del mercato: i consumatori trovano più facilmente i prodotti, i commercianti vendono di più perché le piattaforme pubblicitarie gli permettono di raggiungere un pubblico più ampio attraverso la pubblicità mirata basata sulle preferenze degli utenti, e le piattaforme come Google e Facebook facilitano le transazioni. Tuttavia, il modello comporta alcune sfide nel comprendere come avviene realmente la distribuzione del surplus tra gli stakeholder (utenti, commercianti e piattaforme). Quando gli utenti riescono a trovare rapidamente ciò che cercano, sono più propensi ad effettuare un acquisto. Tuttavia, alcune ricerche suggeriscono che questa stessa pubblicità comportamentale può portare a scelte svantaggiose, inducendo i consumatori ad acquistare prodotti di qualità inferiore a prezzi più elevati. Questo potrebbe avvenire perché, quando gli utenti trovano rapidamente un articolo, tendono a considerare più vantaggioso l'acquisto immediato piuttosto che cercare ulteriormente online, dove potrebbero trovare alternative più convenienti. In altre parole, il risparmio di tempo offerto dalla pubblicità mirata può sovrastare la ricerca di prezzi più competitivi, evidenziando come i costi di ricerca siano un fattore predominante nelle decisioni di acquisto. Le piattaforme digitali riducono i costi di ricerca e favoriscono la competizione, ma allo stesso tempo creano nuove opportunità per editori e contenuti innovativi, aumentando la concorrenza per catturare l'attenzione dei consumatori. In questo contesto, i dati degli utenti vengono centralizzati nelle mani delle piattaforme, che ne traggono i maggiori benefici. Sebbene la pubblicità online aumenti il tasso di conversione per i commercianti, questi devono investire maggiormente rispetto ai metodi tradizionali per non perdere quote di mercato.

2.4 Sistema di tracciamento dei dati

Uno studio tedesco svolto da Hannes Ullrich, Jonas Hannane, Christian Peukert, Luis Aguiar e Tomaso Duso esplora come i metodi di tracciamento dei dati dei consumatori da parte dei principali attori del mercato (come Google, Facebook e Amazon) contribuisca a differenziare le capacità competitive tra le diverse piattaforme, fornendo dei dati empirici a sostegno della tesi. L'esperimento consiste nell'utilizzo di metodi di machine learning per costruire modelli predittivi sulle caratteristiche demografiche degli utenti attraverso i dati di navigazione raccolti dai tracker dell'azienda di marketing Comscore su vari siti web, e che hanno un forte impatto sulla personalizzazione dei prezzi e su altre strategie di marketing mirato. I tracker sono piccoli programmi o frammenti di codice inseriti nei siti web, che monitorano le attività degli utenti e registrano informazioni su pagine visitate, tempi di permanenza e preferenze di contenuti.

La società registra il comportamento online di 75.000 utenti nel web nell'arco di 12 mesi, raccogliendo informazioni demografiche di base come età, livello di istruzione, posizione

geografica e reddito familiare. In particolare, le informazioni raccolte comprendono il nome del dominio visitato, l'ora e la durata della visita per ogni sito web. Inoltre, per migliorare la precisione delle previsioni sul comportamento degli utenti, i dati di navigazione vengono combinati con l'integrazione di informazioni da HTTPArchive e WhoTracks.me, essenziali per l'analisi. Il primo identifica quali siti web vengono monitorati tramite servizi di tracciamento di terze parti: in particolare, i dati di HTTPArchive coprono circa il 77% di tutti i file registrati e il 12% dei siti di alto livello nella classifica di Comescore; infine, catturano il 99% degli utenti registrati nel set di dati. Inoltre, presuppone che questi tracker siano rimasti stabili nel tempo, facilitando un'analisi comparabile tra il periodo antecedente, di circa 20 anni, e successivo al GDPR. Invece, le informazioni di WhoTracks.me aiutano a confrontare i dati raccolti con un elenco di tracker conosciuti, fornendo dettagli su quali domini monitorano i dati e quali aziende li gestiscono. Questo permette di osservare i 50 tracker più utilizzati e come sono distribuiti sui siti analizzati: Google, Facebook e Amazon emergono come i principali fornitori di tracker, con un numero significativo di domini monitorati. Per valutare l'attività di monitoraggio e l'intensità del traffico su ciascun dominio, l'analisi tiene conto del numero di clic effettuati per visita, che consente di comprendere meglio il comportamento di navigazione e come i tracker monitorino specifici domini in base alle caratteristiche degli utenti e alla frequenza delle visite.

Il campione completo è costituito da 75.150 osservazioni, ovvero gli utenti tracciati, con un totale di 155.247 domini. Si costruiscono i profili demografici dei consumatori attraverso il monitoraggio web delle loro abitudini di navigazione, studiando come questi reagiscono e interagiscono con vari domini web. Infine, si raggruppano i vari utenti in 39 gruppi demografici correlando le caratteristiche in comune.

Per valutare le prestazioni dei modelli di previsione si utilizza la metrica AUC (Area Under the Curve), analizzando l'area sotto la curva ROC (Receiver Operating Characteristic). In particolare, un AUC fino a 0,5 indica una previsione casuale, ovvero il modello non è efficace; mentre un AUC compresa tra 0,5 e 1 indica una previsione accurata, quindi un modello molto efficace.

2.4.1 Risultati

L'accuratezza delle previsioni dipende dal tipo di dati che vengono raccolti e da come vengono successivamente elaborati. In particolare, vengono eseguite le previsioni su quattro distinte variabili demografiche:

- Dimensione della famiglia

- Età
- Presenza di figli
- Censimento Regione

I risultati dell'analisi sulle previsioni di accuratezza realizzate da vari tracker evidenziano che la qualità delle previsioni varia molto in base al tipo di variabile demografica. Ad esempio, sulla prima variabile "dimensione della famiglia" si rileva un AUC di circa 0,53, ciò significa che il modello non è efficace e che nemmeno le grandi piattaforme, come Google e Facebook, ottengono previsioni accurate. Invece, sulla variabile "età" si ha un AUC medio rispettivamente 0,68 e 0,76, il che significa che sono previsioni moderatamente accurate, dunque senz'altro migliori della variabile demografica precedente. Pertanto, quando un modello è poco efficace, ossia quando la qualità dei dati non è adeguata a ottenere risultati significativi, nemmeno le grandi piattaforme, con la loro vasta quantità di dati, riescono a ottenere un vantaggio concreto.

In secondo luogo, per studiare la relazione tra dimensione dei dati e qualità delle previsioni, gli studiosi fanno un'analisi attraverso una griglia 10x10 nella quale inseriscono in maniera casuale diverse percentuali di dati (ad esempio 0%, 10%, fino a 90%), per vedere come cambia l'accuratezza del risultato in base alla quantità di dati utilizzata. Per gestire questo ampio set di dati vengono utilizzati algoritmi avanzati come LightGBM, che simulano ripetutamente l'esperimento utilizzando in input diverse frazioni di dati riportate nella griglia. I risultati dell'esperimento suggeriscono che, anche se si ha accesso ad una grande mole di dati, le performance predittive dipendono dalla qualità intrinseca di questi dati e dalla complessità della previsione, la quale è differente per ogni variabile che si vuole analizzare. Di conseguenza, per fare previsioni accurate non basta avere una grande quantità di dati, ma è fondamentale che questi siano anche pertinenti e informativi per lo specifico obiettivo predittivo. Se i dati raccolti non contengono informazioni rilevanti per la previsione che si sta cercando di fare, non saranno particolarmente utili. Le grandi aziende sono avvantaggiate nell'ottenere risultati migliori grazie all'accesso ad un numero molto alto di domini e utenti, che invece può costituire un ostacolo all'ingresso per i tracker più piccoli. Tuttavia, l'analisi fatta attraverso l'algoritmo di LightGBM dimostra che, benché la qualità della previsione tenda a migliorare con l'aggiunta di dati demografici, l'effetto positivo si riduce progressivamente quando il volume dei dati raccolti dai tracker raggiunge determinate soglie. Tracciando un grafico di distribuzione dell'AUC tra attività di monitoraggio e previsione con i dati attuali, risulta che i miglioramenti nella capacità predittiva non sono lineari: la qualità predittiva aumenta inizialmente con il numero di domini monitorati, ma oltre una determinata soglia l'effetto positivo si stabilizza o diminuisce,

segnalando che oltre questo punto la raccolta di ulteriori dati non apporta un miglioramento significativo.

Inoltre, si evidenzia l'importanza della "complementarità dei dati", ovvero il valore aggiunto che dati diversi tra loro (provenienti da molteplici fonti e tipi di attività) possono portare alle previsioni rispetto ai dati dello stesso tipo. Questo concetto è collegato alla teoria delle rese decrescenti: l'aggiunta di un certo tipo di dato può migliorare le previsioni inizialmente, ma oltre un certo punto i benefici diminuiscono e per migliorare ulteriormente la qualità delle previsioni è necessario integrare dati diversi tra loro.

In conclusione, questo studio suggerisce che le grandi aziende, grazie all'accesso a una quantità enorme di dati, potrebbero avere un vantaggio competitivo, ma avere un grande volume di dati aggiunge valore fino a un certo punto, dopo il quale una quantità eccessiva di dati può avere un impatto marginale sulle prestazioni.

2.4.2 Combinazione di dati clickstream e dati demografici

Nonostante la diminuzione del numero di domini e utenti tracciati a causa dell'aumento delle normative sulla privacy, i grandi tracker di dati (come i servizi di Google o Facebook) hanno ancora il primato sull'accesso diretto a un'enorme quantità di informazioni personali e dati di navigazione. Il loro ruolo centrale nei servizi online gli permette di raccogliere dati su scala molto ampia, spesso in modi che comprendono sia le interazioni dirette degli utenti con i motori di ricerca, sia i loro comportamenti di navigazione, come i link cliccati o il tempo trascorso su determinati contenuti. Facebook, ad esempio, oltre ai contenuti pubblicati dagli utenti, raccoglie informazioni sulle interazioni sociali (chi sono gli amici, quali post vengono commentati, messaggi scambiati) e sul comportamento di navigazione all'interno della piattaforma. Utilizzando anche pixel di tracciamento su siti web di terze parti che collaborano con loro, riescono a monitorare quali pagine vengono visitate e quali contenuti attirano l'attenzione. L'enorme quantità di dati raccolti gli permette di creare profili estremamente dettagliati che includendo modelli di comportamento, preferenze d'acquisto e persino dati predittivi sui bisogni e interessi futuri. Senza dubbio, questo accesso vasto e continuo ai dati è un ostacolo significativo per nuovi concorrenti nel mercato del web tracking, in quanto le risorse e la scala di Google, Facebook e Amazon sono difficili da eguagliare.

Uno studio analizza fino a che punto le grandi piattaforme possano migliorare la loro capacità di previsione, rispetto ai tracker più piccoli, combinando i dati di navigazione o clickstream con quelli demografici. L'obiettivo è capire se, sfruttando tale combinazione, queste piattaforme

sono in grado di determinare con precisione l'età degli utenti, basandosi sui loro comportamenti online e sugli interessi tracciati, e successivamente di raggrupparli in specifiche fasce di età demografiche. Per valutare in modo accurato se l'aggiunta di dati demografici ai dati di clickstream migliorasse l'accuratezza predittiva rispetto all'utilizzo dei soli dati di clickstream, è stato sviluppato il seguente modello di regressione lineare:

$$Y_{i,s,f,p} = \alpha + \beta_1 demographics_p + \gamma X_{i,s,f,p} + \varepsilon_{i,s,f,p}$$

dove:

- la variabile di risultato Y è l'AUC (Area Under the Curve) che riflette la qualità della previsione per il tracker "i", sull'attività "s", attraverso il modello previsionale "p";
- demographics è una variabile binaria che assume valore "1" se è disponibile l'accesso ai dati demografici, "0" altrimenti;
- X indica ulteriori variabili aggiuntive;
- beta rappresenta il contributo dei dati demografici alla qualità della previsione.

I risultati mostrano che l'integrazione dei dati demografici con i dati di clickstream migliora mediamente le prestazioni predittive di 0,07, ottenendo un aumento della qualità di circa l'11% rispetto al caso di utilizzo dei solo dati clickstream. Ciò evidenzia il valore aggiunto che le informazioni demografiche possono offrire nelle attività di previsione. Pertanto, questa analisi è rilevante anche dal punto di vista antitrust: i gatekeeper come Google, Facebook o Amazon, grazie all'accesso esclusivo ai dati demografici, ottengono un vantaggio competitivo significativo che le aziende più piccole, senza tale accesso, non possono colmare facilmente. Per affrontare le disparità nel mercato digitale, entrano in gioco normative europee come il Digital Markets Act (DMA), che mira a limitare il controllo dei dati da parte delle grandi aziende tecnologiche e a promuovere una concorrenza più equa. In particolare, il DMA propone che i gatekeeper come Google e Facebook condividano parte dei dati demografici raccolti con attori più piccoli, come i tracker non dominanti. Questo permetterebbe anche a queste aziende di combinare tali informazioni con i dati derivati dal clickstream, migliorando così la qualità delle loro previsioni e aumentando la competitività. Tuttavia, questo approccio solleva una questione cruciale legata alla protezione dei dati personali. La condivisione dei dati degli utenti tra piattaforme richiederebbe il consenso esplicito dei proprietari dei dati, in linea con il Regolamento Generale sulla Protezione dei Dati (GDPR). Il DMA, dunque, si trova a dover affrontare un equilibrio delicato: garantire che i benefici della condivisione dei dati siano sfruttati per ridurre le disuguaglianze tra attori del mercato, senza violare le normative sulla protezione dei dati e senza mettere a rischio la fiducia degli utenti.

2.5 Economie di scopo

Negli ultimi anni i casi antitrust hanno messo in luce le crescenti preoccupazioni riguardo l'aggregazione di dati complementari raccolti dalle grandi piattaforme digitali. Queste preoccupazioni si accentuano quando le grandi aziende acquisiscono piattaforme più piccole per ottenere accesso a ulteriori set di dati, costruendo così database sempre più vasti e ricchi. Un esempio significativo è rappresentato dall'acquisizione di Instagram da parte di Facebook, che ha permesso a quest'ultima di ampliare enormemente la propria base informativa integrando i dati degli utenti di entrambe le piattaforme. In questo contesto nasce il tema di "economie di scopo", inteso quando per un'azienda è più conveniente economicamente combinare due o più linee di prodotti piuttosto che produrle separatamente. In altre parole, le economie di scopo si verificano quando un'azienda possiede una risorsa che non viene sfruttata completamente e può essere utilizzata per generare più prodotti o servizi. Così, la capacità inutilizzata di un input viene ottimizzata creando output multipli. È importante distinguere le economie di scopo dalle economie di scala: le prime si riferiscono alla capacità di diversificare l'offerta, ampliando la gamma di prodotti o servizi realizzati simultaneamente; al contrario, le economie di scala si riferiscono all'aumento della produzione di un unico prodotto, ottenendo vantaggi di costo su una maggiore quantità dello stesso output. Inoltre, mentre le risorse materiali rivali, come macchine e attrezzature fisiche, non possono essere impiegate per produrre più output contemporaneamente, lo stesso non vale per i beni immateriali e non rivali, come i dati o i software. Quando si lavora con un set di dati, fare economie di scopo significa arricchire il modello con nuove variabili indipendenti che possano migliorare l'accuratezza delle previsioni. Tuttavia, è fondamentale che queste nuove variabili siano selezionate con criterio: non devono essere semplici "repliche" di variabili già presenti, cioè altamente correlate al punto da non aggiungere informazioni utili, né del tutto slegate dalle variabili esistenti, per evitare che generino solo "rumore" anziché apportare vantaggi significativi. L'obiettivo è creare un insieme di variabili che si completino a vicenda, in modo che ciascuna contribuisca realmente al valore informativo del modello, migliorando così la capacità di previsione senza ridondanze né elementi superflui. Tuttavia, estrarre informazioni aggiuntive da dati grezzi può comportare dei costi in termini di tempo e risorse maggiori rispetto ai benefici dell'analisi aggiuntiva, rendendo non conveniente l'estrazione di tutte le informazioni possibili. Per questo motivo, bisogna considerare modelli di apprendimento economico che aiutino a selezionare solo quelle combinazioni di dati e quelle analisi che aggiungono effettivo valore, ovvero focalizzandosi su quelle che possono realmente migliorare la qualità delle previsioni e delle strategie decisionali. In conclusione, si realizzano economie di scopo nell'analisi esplorativa dei dati (ESDA) quando l'integrazione di più dataset permette di ottenere un livello informativo più alto, ovvero

un'accuratezza e una ricchezza di informazioni superiori rispetto a quelle ottenibili dai singoli dataset presi separatamente.

I dati rappresentano un asset strategico in grado di generare contemporaneamente output diversi, favorendo l'ottimizzazione delle economie di scopo. Una caratteristica peculiare è la possibilità per più soggetti di utilizzare lo stesso set di dati per scopi differenti, senza compromettere la funzionalità originaria o l'accesso da parte di chi li ha raccolti. Ciò è possibile grazie al fatto che la loro raccolta implica solitamente un costo fisso iniziale, ma una volta acquisiti, possono essere riutilizzati più volte con costi marginali quasi nulli per la duplicazione. Questa caratteristica ha portato all'idea diffusa che la condivisione dei dati, una volta acquisiti, apporterebbe vantaggi alla società, poiché il loro utilizzo per molteplici finalità a costo zero potrebbe aumentare i benefici complessivi senza richiedere ulteriori spese di raccolta.

Tuttavia, sebbene un accesso più aperto possa portare benefici alla società, Palfrey e Gasser (2012) sottolineano il fatto che non tutti i soggetti coinvolti ne potrebbero trarre vantaggio. Da un lato, le persone sono spesso contrarie alla diffusione indiscriminata dei loro dati personali; dall'altro, le aziende preferiscono mantenere riservati i loro dati commerciali per proteggere il vantaggio competitivo. In altre parole, se i dati venissero condivisi liberamente, il loro valore economico potrebbe ridursi notevolmente e scendere al di sotto del costo iniziale per produrli. Si verrebbe a creare, così, un divario tra il valore sociale (i benefici per la collettività) e il valore privato (i vantaggi per chi possiede i dati). I regolatori potrebbero obbligare le aziende a condividere i dati se ritenessero che i vantaggi sociali siano sufficientemente superiori alle perdite private, e potrebbero prevedere compensazioni per chi subisce perdite economiche dalla condivisione obbligatoria.

2.5.1 Teoria ESDA

L'area dell'ESDA (Economie di Scopo nei Dati Aggregati) è ancora poco esplorata nella letteratura scientifica e, di conseguenza, non ci sono delle rilevanti prove empiriche a testimonianza di effettivi vantaggi. Comunque sia, ci sono stati diversi studi portati avanti con lo scopo di comprendere i benefici derivanti dall'aggregazione dei dati in termini di precisione delle previsioni di vendita.

Uno studio empirico intitolato "Scope Economies in Data" di Bruno Carballa-Smichowski, Néstor Duch-Brown, Seyit Hocuk, Pradeep Kumar, Bertin Martens, Joris Mulder, Patricia Prufer, analizza i vantaggi economici derivanti dall'integrazione di dataset con dati eterogenei,

che rendono l'aggregazione una strategia molto conveniente permettendo di rafforzare la posizione di mercato per le aziende di grandi dimensioni, riducendo le possibilità per le aziende più piccole di competere efficacemente. Lo studio, in particolare, si concentra su come migliorare l'accuratezza delle previsioni mediche utilizzando diversi set di dati diversificati riguardanti dati sanitari e no, raccolti a livello individuale. Tutto ciò avviene grazie alla possibilità di condividere le cartelle cliniche elettroniche dei pazienti, facilitando la creazione di modelli di previsione avanzati, che riescono a stimare la salute futura degli individui utilizzando una varietà di dati personali. A tal proposito è stato molto rilevante il contesto del COVID-19 che ha ulteriormente evidenziato il valore dell'integrazione di dati sanitari e non sanitari per elaborare strategie di contenimento dell'infezione, dimostrando che l'analisi di set di dati complementari può avere un impatto significativo sul benessere pubblico. Questa crescente raccolta e analisi di dati sanitari e personali comporta, tuttavia, questioni rilevanti sulla privacy e sulla protezione dei dati, poiché le informazioni sanitarie sono dati sensibili che richiedono un'attenzione rigorosa alla sicurezza.

Per l'analisi empirica sono stati presi in considerazione due fonti di dati provenienti dai Paesi Bassi:

- Un set di dati dall'indagine socioeconomica Longitudinal Internet Studies for the Social Sciences (LISS) il quale comprende 5000 famiglie olandesi, con un totale di circa 7.500 individui, che completano sondaggi online mensili a partire dal 2007 su un'ampia gamma di argomenti, come salute, personalità, reddito, patrimonio e così via. In particolare, i sondaggi comprendono 127 domande di indagine sul tema salute, di natura soggettiva. In totale, dalle risposte dei sondaggi sono state estratte in questo set di dati 2.007 variabili predittive e un totale di 89.611 osservazioni.
- Un set di microdati a livello individuale per l'intera popolazione, raccolti dall'Ufficio centrale di statistica olandese (CBS), su un'ampia gamma di argomenti a partire dal 2006. In particolare, sono stati utilizzati i dati sulle prescrizioni di farmaci provenienti dalla sezione "Salute e benessere" del database, che copre l'intera popolazione dei Paesi Bassi. Questo processo ha prodotto 187 variabili predittive aggiuntive utili per l'analisi.

Tra i membri delle due fonti c'è una corrispondenza univoca solo per coloro che hanno acconsentito all'accoppiamento di questi dati tramite il loro ID. Per quanto riguarda la qualità della copertura tra i due dataset, si è rilevata una sovrapposizione del 92% tra le persone nel dataset CBS sui medicinali e i partecipanti al panel LISS. Invece, sono stati esclusi dall'analisi quei dati la cui sovrapposizione di individui tra i due set era troppo bassa per garantire coerenza

e completezza allo studio. La maggior parte dei dati è stata raccolta nel periodo di cinque anni compreso tra il 2015 e il 2019: si è scelto di escludere il 2020 poiché lo si è considerato non rappresentativo a causa dell'impatto della pandemia di COVID-19. Inoltre, non è stato ritenuto utile estendere la raccolta a periodi precedenti, poiché ciò avrebbe aumentato il numero di osservazioni per ciascun partecipante (ossia le rilevazioni nel tempo), ma non il numero complessivo di partecipanti stessi. Dopo la raccolta di tutte le variabili con le rispettive osservazioni, il primo set di dati è stato "ripulito" da tutte le variabili che contenevano informazioni non correlate o inutilizzabili. Infatti, si è passati ad un set finale di 512 variabili; mentre, il secondo set è rimasto invariato con 187 variabili. Dopo di che, in fase di elaborazione sono stati calcolati dei coefficienti, da -1 a 1, per determinare il livello di correlazione tra le variabili del dataset, e per capire quelle fortemente correlate che dovevano essere eliminate fissando arbitrariamente delle soglie di correlazione (0,5, 0,7, 0,9) per confrontare i diversi risultati al variare del numero di variabili predittive utilizzate. Tutto ciò per migliorare le prestazioni del modello predittivo riducendo ridondanze e distorsioni che non aggiungono nuove informazioni. Alla fine della pulizia e trasformazione dei dati, il dataset aggregato conteneva 807 variabili e 22.792 osservazioni.

La scelta della variabile di esito può influire notevolmente sui risultati poiché alcune rispondono più efficacemente all'aggregazione dei dati rispetto ad altre. Le variabili di tipo oggettivo, essendo più facilmente misurabili, tendono a richiedere un numero ridotto di dati per ottenere previsioni accurate. Al contrario, le variabili soggettive richiedono un insieme di dati più ampio e diversificato, che includa sia informazioni sanitarie sia dettagli socioeconomici per rappresentare con precisione la complessità dell'esito considerato. Ciò rende le variabili soggettive più adatte alle previsioni di machine learning rispetto alle variabili oggettive, le quali possono essere previste con precisione utilizzando l'analisi di regressione. Così, sono state scelte due variabili soggettive in ambito sanitario: salute percepita e disabilità funzionale, ognuna delle quali prevede per i rispondenti una scala da 1 a 5, raggruppando poi le risposte in diverse categorie.

L'addestramento dell'algoritmo di machine learning inizia utilizzando una piccola percentuale delle variabili predittive disponibili e andando ad incrementare progressivamente. L'obiettivo è valutare come l'aggiunta di nuove variabili influisca sulle prestazioni del modello nella previsione dei risultati sanitari. Le variabili impiegate durante l'addestramento dell'algoritmo appartengono al set di training, mentre quelle utilizzate successivamente per effettuare le previsioni fanno parte del set di test. Tuttavia, per aumentare la robustezza delle previsioni ed evitare che i risultati dipendano da una singola suddivisione, si usa la convalida incrociata k-

fold. Con questa tecnica, i dati vengono mescolati e divisi in k gruppi o "fold" dove ogni fold viene usato a turno come set di test, mentre i restanti fungono da training set, per permettere al modello di essere addestrato e valutato su tutte le parti del dataset migliorando l'accuratezza e la stabilità delle metriche. Le metriche utilizzate per valutare le prestazioni del modello sono:

- Precisione per ogni classe, che misura la correttezza delle predizioni fatte per ciascuna categoria.
- Richiamo, ovvero la capacità del modello di identificare correttamente tutti gli esempi positivi di una classe, garantendo che nessun caso rilevante venga trascurato.
- Punteggio F1, che rappresenta la media armonica tra precisione e richiamo.

Quest'ultima metrica, il punteggio F1, è particolarmente adatta per dataset sbilanciati, in cui alcune classi sono significativamente meno rappresentative di altre, e fornisce una valutazione complessiva del modello bilanciando accuratezza e completezza nelle predizioni.

2.5.2 Economie di scopo nell'aggregazione dei dati

I risultati preliminari dell'analisi dell'ESDA vengono successivamente rappresentati graficamente, in particolare vengono analizzate le prestazioni predittive, considerando come livello di correlazione 0,7 e il periodo 2018-2019, utilizzando due algoritmi di apprendimento automatico: Random Forest (RF) e Logistic Regression (LR). L'obbiettivo principale è verificare la specifica ipotesi legata al concetto di ESDA, ovvero come l'accuratezza predittiva migliori all'incrementare del numero di variabili esplicative incluse nel modello, mantenendo costante il numero di osservazioni. I grafici mostrano come il punteggio F1, che misura l'accuratezza del modello, aumenti man mano che si usano più variabili esplicative. Tuttavia, questo aumento non è sempre graduale, poiché l'accuratezza può migliorare significativamente con l'aggiunta di variabili utili, ma può anche diminuire se vengono aggiunte variabili irrilevanti, aumentando la complessità del modello e il rumore.

Con l'obbiettivo di testare la robustezza dei risultati, vengono stimati tre modelli statistici:

$$Modello\ 1:\ \log(score) = \beta_0 + \beta_1 \log(VarPerc) + \beta_2 RF + \beta_3 DIS + \beta_4 \varepsilon_i$$

$$Modello\ 2:\ \log(score) = \delta_0 + (\delta_1 RF + \delta_2 DIS + \delta_3 S)\log(VarPerc) + \varepsilon_i$$

$$Modello\ 3:\ \log(score) = \alpha_{0+}\alpha_1 \log(VarPerc) + \alpha_2 RF + \alpha_3 DIS + \alpha_4 S + (\alpha_5 RF + \alpha_6 DIS + \alpha_7 S)\log(VarPerc) + \varepsilon_i$$

dove:

- log (score) è il logaritmo naturale del punteggio F1;
- log (VarPerc) è il logaritmo naturale della percentuale di variabili incluse in esso;
- RF è una variabile dummy che indica il tipo di algoritmo utilizzato;
- DIS è una variabile dummy che indica il tipo di variabile prevista;
- S è una variabile dummy che indica il periodo di analisi.

L'obbiettivo dei tre modelli è vedere come alcune caratteristiche specifiche dei modelli di machine learning, come il tipo di algoritmo utilizzato, la variabile di previsione e il periodo di analisi, possono influenzare l'accuratezza predittiva. Il primo modello valuta l'impatto di ogni singola caratteristica sull'accuratezza media delle previsioni, rappresentata dal livello dei grafici. Il secondo modello analizza come ciascuna caratteristica contribuisca in modo diverso all'entità del miglioramento dell'accuratezza predittiva, rappresentata dalla pendenza dei grafici, che riflette il tasso di incremento dell'accuratezza al crescere della percentuale di variabili esplicative. Infine, il terzo modello combina gli effetti medi e di pendenza per una visione più completa.

Nel Modello 1, l'aumento dell'1% delle variabili totali porta ad un incremento della precisione predittiva, rappresentata dal punteggio F1, da 0,087% a 0,132% mediamente. Invece, il confronto tra le dummy dell'intercetta, il coefficiente di log (VarPerc) e la costante evidenzia che le altre caratteristiche del modello, come il periodo di analisi e il tipo di algoritmo utilizzato, sebbene statisticamente significative, hanno un impatto molto inferiore. Questi risultati dimostrano che la percentuale di variabili esplicative è il principale determinante del livello medio di accuratezza predittiva. Nel Modello 2, invece, si evidenzia che l'algoritmo utilizzato e il periodo di tempo considerato influenzano la pendenza della relazione tra percentuale di variabili esplicative e accuratezza predittiva: in altre parole, sebbene l'aggiunta di variabili esplicative migliori l'accuratezza predittiva, l'entità di questo miglioramento dipende dalle altre caratteristiche del modello.

Tuttavia, lo studio dimostra che all'aumentare della percentuale di variabili esplicative incluse nell'analisi di previsione, l'accuratezza del modello aumenta, ma ad un ritmo sempre più lento. Questo suggerisce l'esistenza di rendimenti decrescenti nell'aggregazione dei dati, i quali implicano che, superata una certa soglia di variabili incluse, l'aggiunta di ulteriori variabili apporta benefici marginali sempre minori. In sostanza, la relazione tra la percentuale di variabili incluse nel modello e l'accuratezza della previsione è a forma di S.

2.5.3 Complementarità tra variabili

Per testare se e come una maggiore complementarità tra le variabili migliori l'accuratezza predittiva nel modello ESDA, si stabiliscono diverse soglie di correlazione (0,5, 0,7 e 0,9) tra le variabili esplicative e per ognuna di esse si stima un'equazione. Le variabili con un coefficiente di correlazione superiore alla soglia specificata vengono **eliminate** dal set di dati. In particolare, un valore basso della soglia implica una maggiore eliminazione di variabili altamente correlate, lasciando solo variabili più complementari. Viceversa, un valore alto della soglia permette l'inclusione di variabili più simili tra loro, riducendo la complementarità complessiva del set di dati. Si dimostra che maggiore è la complementarità, più significativi sono i miglioramenti nell'accuratezza predittiva all'aumentare del numero di variabili esplicative. Pertanto, quando la soglia di correlazione viene abbassata vengono eliminate più variabili correlate, riducendo il numero totale di variabili disponibili. Tale riduzione può limitare la capacità del modello di sfruttare pienamente le informazioni nei dati; di conseguenza, anche se la complementarità aumenta, il beneficio complessivo in termini di accuratezza potrebbe non crescere in modo proporzionale. Infatti, si nota che:

- una diminuzione da 0,7 a 0,5 della soglia di correlazione porta a una riduzione dell'ESDA di 0,043 punti;
- una diminuzione da 0,9 a 0,7 della soglia di correlazione porta a una riduzione dell'ESDA di solo 0,002 punti.

Questo suggerisce che aumentare la complementarità delle variabili è più efficace quando il livello iniziale di complementarità è basso, ovvero quando la soglia di correlazione tra variabili è, ad esempio, a 0,9. Quando, invece, la soglia è già moderatamente bassa, come 0,7, abbassarla ulteriormente elimina altre variabili, ma i benefici di una maggiore complementarità diventano meno significativi.

Le economie di scopo derivanti dall'aggregazione di dati complementari giustificano la creazione di ampi pool di dati, come proposto nella strategia europea per i dati, che possono generare un maggiore benessere sociale. Tuttavia, sorgerebbero questioni di equità quando tali pool sono sotto il controllo esclusivo di operatori privati, rischiando di far nascere monopoli e asimmetrie informative. Per bilanciare questi aspetti, l'accesso condiviso ai pool di dati può rappresentare una soluzione, permettendo di realizzare economie di scala senza cedere il controllo esclusivo dei dati.

2.6 Confronto tra motori di ricerca

Sorge il dubbio su come motori di ricerca come Google e Facebook riescano a fornire risultati migliori: è grazie all'efficienza dei loro algoritmi o al fatto che hanno accesso a una cronologia di dati storici più ampia, che permette loro di raccogliere informazioni più dettagliate? È stato osservato che, man mano che cresce il numero di utenti che utilizzano un motore di ricerca, questi ultimi hanno la possibilità di migliorare nel tempo grazie alla grande quantità di dati storici accumulati. Lo scopo dell'analisi è, dunque, capire se le aziende che dominano questo mercato beneficiano di un vantaggio competitivo dovuto all'accesso a un volume di dati decisamente maggiore rispetto alle nuove realtà che cercano di entrare nel settore, o se semplicemente dispongono di algoritmi migliori per l'analisi dei dati. L'applicazione delle regolamentazioni antitrust dipende molto da tale questione, poiché, se il successo futuro di una piattaforma è davvero solo una conseguenza dell'efficienza degli algoritmi utilizzati, e non di un accesso privilegiato ai dati degli utenti, ci sarebbero meno motivi per applicare regolamentazioni. Se, invece, non dovesse trattarsi dell'algoritmo, si solleverebbero preoccupazioni relative al fatto che le grandi aziende possiedono un accesso esclusivo a certi dati, creando disuguaglianze e ostacolando l'innovazione, poiché non hanno incentivi a migliorarsi continuamente.

Uno studio di Tobias J. Klein, Madina Kurmangaliyeva, Jens Prufer Patricia Prufer

confronta la qualità dei risultati di ricerca generati da grandi motori di ricerca, come Google e Bing, con quelli di un motore di ricerca meno rilevante, in questo caso Cliqz, per capire come la quantità di dati storici influisca sulla qualità dei risultati. Attraverso il software Human Web integrato nel browser di Cliqz, che consente di accedere alle ricerche effettuate dagli utenti sul motore di ricerca, Cliqz ha selezionato tutte le query digitate dagli utenti nel periodo settimanale che intercorre tra il 20 ed il 26 aprile 2020. Questo processo ha portato alla creazione di 5 sottogruppi, ognuno rappresentante una specifica percentuale del totale delle query (0,2%, 1%, 5%, 25%, 75%). Infine, da ogni raggruppamento sono state estratte 1000 query arrivando ad un numero complessivo di 5.000. Successivamente, l'esperimento ha simulato diversi scenari variando la quantità di dati storici degli utenti a disposizione del motore di ricerca. Si è passati da una quantità limitata a una più ampia, per analizzare come la qualità e la rilevanza dei risultati cambiassero in relazione alla quantità di informazioni disponibili. Per ogni query selezionata, sono stati testati 12 livelli di accesso ai dati, variando la disponibilità dal 100% fino allo 0,1%. In totale sono stati analizzati 60.000 set di risultati generati da Cliqz e 5.000 set provenienti rispettivamente da Google e Bing; questo approccio ha permesso di analizzare l'impatto della

riduzione dei dati sull'efficacia delle risposte del motore di ricerca. Successivamente, sono state scelte casualmente 500 query, dalle 1.000 originali, in modo stratificato dai diversi livelli di popolarità. Infine, eliminando le query meno popolari e i contenuti inappropriati si è arrivati ad un totale di 493 query da sottoporre a valutazione umana. I risultati di queste vengono confrontati con quelli di Google e Bing per le stesse query grazie a valutatori che utilizzano una scala Likert da 0 a 7 per misurare la qualità e la pertinenza dei risultati, consentendo una valutazione comparativa dell'efficacia dei vari motori di ricerca. Per garantire l'imparzialità della valutazione, sono stati creati set di risultati misti in cui i valutatori non sapevano quale motore di ricerca avesse prodotto ciascun risultato. Questa tecnica consente di analizzare le loro preferenze senza influenze esterne, rendendo particolarmente interessante vedere quali risultati selezionassero come "migliore" e "secondo migliore" all'interno di questi set. Per facilitare l'analisi e ridurre il carico sui valutatori umani, ci si è concentrati solo sui primi cinque risultati di ciascun motore di ricerca, in quanto studi precedenti hanno dimostrato che gli utenti tendono a visualizzare solo le prime posizioni dell'elenco dei risultati. Successivamente, è stato esaminato quale motore di ricerca producesse il miglior risultato con maggiore frequenza, ovvero valutare la percentuale di volte in cui il risultato scelto come "migliore" all'interno del set misto proveniva da uno specifico motore di ricerca. Dai risultati emerge chiaramente che, per qualità delle risposte, Google si posiziona al primo posto, seguito da Bing e, con una differenza significativa, da Cliqz. Per approfondire le possibili cause di queste differenze di prestazioni, le query sono state suddivise in cinque categorie in base alla loro popolarità, consentendo di verificare se la qualità dei risultati dipenda dalla frequenza delle query: Se le differenze emergessero principalmente per le query meno comuni, ciò potrebbe indicare che i dati disponibili giocano un ruolo determinante. I risultati mostrano che, riducendo la quantità di dati ai quali Cliqz poteva avere accesso, diventa più difficile per il motore di ricerca restituire dei risultati validi. Invece, con accesso completo ai dati, Cliqz non è riuscito a fornire alcun risultato soltanto per circa il 3% delle query. In conclusione, per le query più comuni la qualità dei risultati di Cliqz è comparabile a quella di Google e Bing; mentre, invece, per le query meno popolari, la qualità complessiva dei risultati di Cliqz cala nettamente. Questo indica che il problema non risiede nell'algoritmo di Cliqz in sé, ma piuttosto nella quantità limitata di dati disponibili per queste query meno frequenti. Dunque, si può concludere che il divario di prestazioni tra i motori di ricerca è in gran parte dovuto alla disponibilità di dati: Google e Bing, che hanno un maggior numero di utenti e quindi un volume di dati più ampio anche per le query rare, possono affinare meglio i risultati rispetto a Cliqz. In particolare, si nota che per le query più popolari Cliqz riesce a ottenere buoni risultati di ricerca utilizzando solo il 20% dei

dati disponibili, supportando la teoria dell'apprendimento statistico secondo cui oltre un certo punto l'aumento della dimensione del set di dati genera rendimenti decrescenti sulle prestazioni predittive. Viceversa, nelle categorie di query meno popolari la qualità dei risultati di ricerca di Cliqz è molto inferiore e le curve di prestazione continuano a crescere anche quando si utilizza il 100% dei dati disponibili. Questo suggerisce che per le query rare, Cliqz potrebbe migliorare notevolmente la qualità dei suoi risultati se avesse accesso ad un numero maggiore di dati.

Questo scenario apre la possibilità di regolare la condivisione obbligatoria dei dati tra piattaforme di ricerca per favorire un equilibrio competitivo. Dato che i dati sono non-rivali, ovvero possono essere usati da più operatori senza esaurirsi, questa condivisione non danneggerebbe Google, ma aiuterebbe nuovi concorrenti a migliorare la qualità dei risultati di ricerca, in particolare per le query meno comuni, con vantaggi per gli utenti finali.

2.6 Sistema di targetizzazione dei dati

Attraverso l'accesso alla cronologia di navigazione degli utenti, una piattaforma può osservare le valutazioni espresse su vari elementi, come prodotti, film o libri. Questo riscontro diretto viene utilizzato per generare raccomandazioni personalizzate per nuovi utenti, con l'obiettivo di suggerirgli un elemento "target" che non hanno ancora provato, basandosi sulle preferenze di utenti passati. In questo modo il sistema apprende le correlazioni tra le valutazioni e stima la probabilità che il nuovo utente apprezzi ciascun elemento target.

Nel documento di Gunhaeng e Lee Julian Wright viene descritto il modello di un sistema di raccomandazione, in cui una piattaforma seleziona e suggerisce articoli a ciascun utente in modo mirato, basandosi su dati già disponibili come le valutazioni fornite su articoli già provati dagli stessi utenti. Quando un utente riceve una raccomandazione e sceglie di provare l'articolo suggerito, il risultato può essere positivo o negativo, generando così un "guadagno" diverso sia per l'utente che per la piattaforma, a seconda dell'esperienza dell'utente con quell'articolo. L'obiettivo della piattaforma è stimare la probabilità che ogni articolo ottenga una reazione positiva dall'utente; in questo modo, può selezionare l'articolo con la maggiore probabilità di piacere, massimizzando il valore della raccomandazione per entrambe le parti.

Durante l'analisi viene creata una matrice composta di articoli totali disponibili e gli utenti che forniscono le valutazioni sugli articoli, e ogni elemento della matrice corrisponde alla

valutazione che un utente specifico fa su un determinato articolo. Dalla valutazione di ogni utente, la piattaforma apprende le correlazioni tra le valutazioni degli articoli, in modo da poter prevedere le preferenze dell'utente target sugli articoli target e raccomandare quell'articolo. L'obbiettivo dello studio è capire come questo miglioramento della precisione delle previsioni apporti benefici, o danni, agli utenti. In base alla quantità e qualità dei dati disponibili, una piattaforma può scegliere tra un sistema di raccomandazione personalizzato o generico. La differenza principale risiede nell'approccio ai dati e nel livello di personalizzazione offerto:

- Sistema personalizzato: sfrutta la cronologia personale dell'utente target insieme alle correlazioni apprese dai dati di altri utenti. Questo consente di generare suggerimenti altamente mirati e individualizzati, adattandosi alle preferenze specifiche dell'utente.
- Sistema generico: non si basa su informazioni personalizzate, ma utilizza le informazioni aggregate dai dati di tutti gli utenti, come la valutazione media degli articoli o la loro popolarità generale.

La raccomandazione generica rappresenta una soluzione semplice e pratica in contesti con pochi dati personalizzati o per nuovi utenti. Per ottenere suggerimenti più mirati è fondamentale, invece, segmentare gli utenti in sottogruppi specifici e personalizzare le raccomandazioni in base alle loro caratteristiche. Questo approccio consente di individuare i gruppi con maggiore probabilità di apprezzare un determinato articolo, ottimizzando così l'efficacia delle raccomandazioni e massimizzare il valore per ciascun gruppo.

In particolare, si evidenzia che la personalizzazione risulta maggiormente efficace quando esiste una forte correlazione tra l'oggetto che si intende raccomandare e un articolo passato che l'utente ha già utilizzato, chiamato elemento condizionate. Tuttavia, esistono alcuni casi in cui la personalizzazione delle raccomandazioni può non solo essere inefficace, ma addirittura dannosa per gli utenti. Questo accade quando la piattaforma non imposta correttamente la soglia ottimale per ogni utente, ovvero quando non tiene conto in modo adeguato degli interessi e delle preferenze individuali. Se la soglia non è scelta con attenzione, il sistema potrebbe fare raccomandazioni che non corrispondono ai veri gusti dell'utente, riducendone così l'efficacia e peggiorando l'esperienza complessiva. Ad esempio, se un elemento viene sempre raccomandato, ovvero ha una soglia di correlazione troppo bassa, o mai raccomandato, cioè quando ha una soglia di correlazione troppo alta, un ulteriore grado di personalizzazione non ha alcun impatto. Questo perché il comportamento del sistema non cambia indipendentemente dalle nuove informazioni introdotte. Quindi, l'elemento condizionante agisce come un

indicatore chiave per prevedere l'esperienza dell'utente con l'elemento target: se un utente ha apprezzato in passato l'elemento condizionante, il sistema può stimare con alta probabilità che gradirà anche l'elemento target raccomandato.

Entrambi i sistemi personalizzato e generico utilizzano il valore ex-post dei dati, ossia basano le raccomandazioni sulle osservazioni già disponibili nella cronologia e nelle valutazioni raccolte dagli utenti. In alternativa, i dati possono essere analizzati in base al loro valore ex-ante, ossia al potenziale valore che possiedono prima di essere osservati, tenendo conto di tutte le possibili configurazioni e ponderandole secondo le probabilità effettive che le sottendono. Questo approccio offre una valutazione complessiva e anticipatoria del valore dei dati, indipendente dalla loro specifica osservazione. Ovviamente c'è il rischio di fare delle previsioni sbagliate, le quali possono prevedere o falsi positivi, che si verificano quando il sistema consiglia un elemento che l'utente non apprezza, o da falsi negativi, quando invece il sistema non consiglia un elemento che l'utente avrebbe apprezzato. Pertanto, viene stimato un limite inferiore del tasso di errore di previsione, ovvero l'errore che non può scendere al di sotto di una certa soglia, che fornisce una valutazione teorica della qualità del sistema.

Successivamente, i dati vengono suddivisi in tre funzioni, ciascuna delle quali contribuisce in modo specifico al valore del sistema di raccomandazione e al benessere degli utenti.

- Personalizzazione: grazie alle reazioni raccolte su una vasta base di utenti, il sistema raggruppa questi ultimi considerando interessi simili e creando segmenti più mirati.

- Selezione: basandosi sulle preferenze apprese, il sistema individua l'elemento con la probabilità più alta di piacere a ciascun utente all'interno del suo gruppo e lo propone come migliore suggerimento.
- Screening: il sistema filtra gli elementi non adatti ai gusti del gruppo, riducendo la possibilità che l'utente riceva raccomandazioni irrilevanti.

Ad ognuna delle tre funzioni viene assegnato un parametro, e vengono fatte delle simulazioni ripetute per ciascuna combinazione di parametri alternativi, in seguito il confronto dei risultati empirici tra i diversi scenari consente di analizzare come diverse configurazioni di parametri nei sistemi di raccomandazione influenzino l'**utilità media stimata** ricevuta dagli utenti. Risolvendo tutti i parametri a sistema, risulta che in assenza di un sistema di raccomandazione, quando gli utenti esplorano gli articoli in modo casuale, l'utilità media ottenuta è pari a 0,148. Con l'introduzione dello screening, l'utilità media aumenta di 0,045, mentre aggiungendo anche

la personalizzazione, l'incremento sale di 0,110. Questo evidenzia che lo screening da solo ha un impatto minore sull'utilità media rispetto alla personalizzazione, che apporta un valore maggiore concentrandosi sulle preferenze individuali dell'utente, piuttosto che su un apprendimento generale basato sugli altri utenti. L'introduzione della selezione, senza le altre due funzioni, comporta un aumento dell'utilità media di 0,194, che sale a 0,209 quando viene combinata con la personalizzazione. Questo dimostra che la selezione migliora significativamente l'utilità media, ampliando il numero di opzioni, tra cui gli utenti possono scegliere. Inoltre, il contributo aggiuntivo della personalizzazione risulta meno rilevante quando la selezione è già presente, suggerendo che gli elementi con valutazioni stabili (sia positive che negative) offrono poco valore aggiunto dalla personalizzazione. Quando i tre parametri vengono combinati, l'utilità media aumenta di 0,223 rispetto all'assenza di un sistema di raccomandazione. L'integrazione delle tre funzioni massimizza il valore del sistema: la selezione amplia le opzioni disponibili, lo screening ne migliora la rilevanza, e la personalizzazione ottimizza l'adattamento alle preferenze individuali dell'utente.

2.7.1 Valore marginale della personalizzazione

Per valutare il surplus generato dall'aumento progressivo del grado di personalizzazione nelle raccomandazioni, viene eseguita una simulazione che prevede la configurazione di 10.000 utenti come gruppo di formazione, utilizzato per "allenare" il sistema attraverso le valutazioni degli utenti, e un gruppo di prova con altri 10.000 utenti, per il quale vengono generate previsioni personalizzate. Il sistema si basa su un elemento target e nove elementi di condizionamento, che influenzano la personalizzazione delle raccomandazioni. L'esperimento incrementa gradualmente il livello di personalizzazione rispettivamente da 1 a 9, definito come la quantità di conoscenza che il sistema acquisisce sulle preferenze degli utenti. Per ogni livello si calcola l'utilità media percepita dagli utenti, confrontandola con l'utilità in assenza di personalizzazione. I risultati mostrano che anche un livello minimo di personalizzazione apporta un beneficio significativo, con un incremento progressivo dell'utilità man mano che il sistema integra ulteriori informazioni, come le correlazioni tra l'elemento target e quelli di condizionamento. In sintesi, l'analisi conferma che la personalizzazione migliora il benessere percepito dagli utenti, con un impatto maggiore al crescere della conoscenza del sistema, contribuendo a massimizzare il valore del sistema di raccomandazioni. Tuttavia, dopo un certo grado di personalizzazione il sistema subisce l'effetto naturale di saturazione: una volta che acquisisce una conoscenza significativa delle preferenze dell'utente, ulteriori informazioni aggiungono valore in misura minore.

Si è visto, inoltre, che esiste una forte interazione tra il grado di personalizzazione, cioè il numero di elementi utilizzati per generare previsioni personalizzate, e la dimensione del gruppo di formazione, ovvero il numero di utenti da cui il sistema apprende. In particolare, è stato osservato che, quando la dimensione del gruppo di formazione è ridotta, incrementare il grado di personalizzazione produce un impatto significativo, e viceversa. Questo indica una forte complementarità tra le due dimensioni nelle fasi iniziali. Tuttavia, quando entrambe le dimensioni raggiungono livelli elevati, i benefici marginali derivanti da ulteriori aumenti diminuiscono. In questo scenario, le due dimensioni tendono a diventare sostitutive, ovvero l'incremento di una può compensare la mancanza di crescita dell'altra senza perdite significative di valore. Per una piattaforma con una base dati iniziale limitata, è strategico investire simultaneamente sia nell'ampliamento del gruppo di formazione (raccogliendo più dati utente) sia nell'aumento del grado di personalizzazione per massimizzare il valore generato. Al contrario, una piattaforma che dispone già di una vasta quantità di dati può concentrarsi su una sola delle due dimensioni, ad esempio migliorando la personalizzazione o espandendo ulteriormente il gruppo di formazione, per ottenere un sistema ancora più efficace.

2.7.2 Valore marginale della selezione

Per misurare l'effetto della selezione, ovvero come l'aumento del numero di articoli target disponibili influisca sul valore marginale percepito dagli utenti, gli autori variano il numero di articoli target da 1 a 15, mantenendo costante il livello di personalizzazione e la soddisfazione di base, e calcolando l'utilità media dell'utente per ciascuna configurazione. I risultati mostrano che l'utilità media aumenta al crescere del numero di articoli, in quanto il sistema ha maggiori possibilità di identificare opzioni migliori per gli utenti. Ma, anche qui, il beneficio marginale diminuisce progressivamente; ovvero, con l'aggiunta di articoli, ogni nuovo articolo apporta un contributo sempre minore all'utilità media complessiva.

2.7.3 Valore marginale dello screening

Il livello di soglia ottimale scelto da una piattaforma di raccomandazione influisce sull'utilità percepita dagli utenti, dove per soglia si intende il livello minimo di certezza che la piattaforma deve raggiungere prima di raccomandare un articolo all'utente. Se una piattaforma ha una soglia bassa significa che raccomanda anche gli articoli che non sono perfettamente allineati con i gusti degli utenti; quando invece ha una soglia alta indica che la piattaforma è molto conservativa, che raccomanda articoli solo se è quasi certa che l'utente li apprezzerà. L'esperimento consiste nel far variare il livello di soglia ottimale della piattaforma tra 0,1 e 0,9

e per ognuna vengono eseguite 1.000 simulazioni, mantenendo costanti tutti gli altri parametri del sistema. Ad esempio, considerando una soglia ottimale pari a 0,5, si confronta l'utilità media per diverse soglie con quella ottenuta usando la soglia ottimale. I risultatati mostrano che ad un qualsiasi allontanamento da quella soglia l'utilità media dell'utente diminuisce, e diminuisce sempre più man mano che ci si allontana da quel valore. Se va al di sotto del valore ottimale, significa che il sistema raccomanda articoli non sufficientemente filtrati, causando insoddisfazione: in generale, una riduzione della soglia di 0,1 rispetto al livello ottimale comporta una riduzione dell'utilità media di circa il 10,9%. Viceversa, quando viene superato il valore ottimale significa che il sistema è troppo selettivo, limitando eccessivamente le raccomandazioni utili: si dimostra che un aumento della soglia di 0,1 rispetto al livello ottimale provoca una riduzione media dell'utilità di circa il 22%. In entrambi i casi, un elevato grado di personalizzazione sarebbe inefficace e non compenserebbe un forte disallineamento della soglia.

2.7.4 Valore marginale di utenti aggiuntivi

Infine, lo studio analizza come la dimensione del training set, ovvero il numero di utenti precedenti i cui dati vengono utilizzati per addestrare il modello) influisce sull'utilità media degli utenti. Vengono effettuate diverse simulazioni al variare del numero di utenti appartenenti al training set, partendo da un gruppo di test composto da un singolo utente, e aggiungendo ad ogni simulazione i dati di 10 nuovi utenti, selezionati casualmente dal gruppo di test, fino ad arrivare ad utilizzare i dati di tutti i 10.000 utenti del gruppo di test iniziale. Misurando l'utilità media degli utenti in base a ciascun gruppo di prova, quindi per ciascun livello di apprendimento, si evince che man mano che il set di addestramento cresce, l'utilità media degli utenti nel gruppo di prova aumenta significativamente, raggiungendo un valore massimo di 0,260 quando vengono utilizzati i dati di tutti i 10.000 utenti. Tale valore rappresenta un miglioramento del 78,87**%** rispetto alla situazione iniziale in cui il sistema disponeva di un singolo utente come training set. Tuttavia, l'impatto marginale positivo dell'integrazione di nuovi dati è decrescente, quindi diminuisce progressivamente man mano che il sistema accumula un numero maggiore di punti informativi.

CAPITOLO 3 – CASO FACEBOOK: INTERSEZIONE TRA PRIVACY E CONCORRENZA

3.1 Origini dell'Antitrust europeo

Per comprendere la logica e la struttura delle leggi Antitrust europee, è fondamentale risalire alle loro origini radicate nel sistema giuridico statunitense. Il termine "trust" nacque negli Stati Uniti nel XIX secolo, riferendosi a pratiche e azioni utilizzate per aggirare il divieto di accordi restrittivi della concorrenza; con il tempo, il termine ha acquisito una connotazione principalmente economica. La regolamentazione Antitrust negli Stati Uniti è stata fortemente influenzata dal pensiero della Scuola di Chicago, il cui obiettivo centrale era l'ottimizzazione del benessere collettivo, includendo sia i consumatori che i produttori. Secondo questa visione, le pratiche aziendali non venivano considerate illegali se contribuivano a migliorare il benessere economico nel complesso, anche se ciò poteva avvenire a scapito del benessere di alcuni consumatori o aziende. La Scuola di Chicago sosteneva che un mercato libero, senza le eccessive regolamentazioni, fosse in grado di autoregolarsi per garantire una concorrenza spontanea e dinamica. Secondo questa visione, la concorrenza tra imprese avrebbe spinto le aziende a migliorare i loro prodotti e servizi con un buon rapporto qualità-prezzo. Tuttavia, nella fase post-chicago si è realizzato il fatto che il libero mercato non aveva effettivamente tutti gli strumenti necessari per limitare l'inefficienza economica ed evitare comportamenti anticoncorrenziali. La prima legge a tutela del mercato fu emanata nel 1890, lo Sherman Act, che sanzionava sia l'effettivo abuso di posizione dominante quando si verificava, sia l'intenzione preventiva di raggiungere tale obiettivo, ossia qualsiasi comportamento che rivelasse le intenzioni dell'azienda di ottenere un vantaggio competitivo ingiusto. Questa legge era stata formulata in termini molto generali, con l'intento di essere applicata caso per caso, e successivamente è stata arricchita e perfezionata con l'introduzione di ulteriori normative. La normativa antitrust europea risale al 1957 ed è contenuta prevalentemente nei Trattati CEE e CECA, con le successive integrazioni ed amplificazioni. Gli articoli 81 e 82 del Regolamento vietano specificamente le intese e le pratiche concordate, considerate illegali a prescindere, per il potenziale danno che potrebbero causare; invece, il semplice possesso di una posizione dominante non è vietato, ma lo diventa solo nel caso in cui viene accertato che tale posizione è stata abusata per ottenere indebiti vantaggi nel mercato. Nonostante le medesime origini, le normative americane e comunitarie si distinsero successivamente per il contesto politico ed economico che dominava il rispettivo territorio nel periodo specifico, il quale determinava una scelta piuttosto che un'altra. Infatti, la legislazione concorrenziale americana è nata in un contesto di integrazione dei mercati federali in uno spazio economico unitario, in cui lo scopo

principale era realizzare processi concorrenziali in grado di mantenere il potere di mercato. Invece, la Comunità Europea alla fine degli anni 50 era ancora fortemente frammentata e lontana dal raggiungimento di un mercato unico. Da ciò si comprende come le legislazioni Antitrust perseguivano obbiettivi diversi nei due continenti: se negli Stati Uniti veniva vista come uno strumento per impedire accordi che minacciavano la concorrenza, mentre in Europa era lo strumento di integrazione fra stati. Nella cultura statunitense, con l'influenza della Scuola di Chicago, vengono ammesse e giustificate imprese di grandi dimensioni se la loro presenza contribuisce all'aumento del benessere collettivo. Invece, in Europa le aziende che esercitano un controllo su porzioni più rilevanti del mercato vengono automaticamente considerate come un ostacolo per l'abbattimento delle barriere. Un'altra differenza significativa nelle legislazioni Antitrust è che negli Stati Uniti queste variano tra i singoli Stati e non sono necessariamente coordinate con la normativa federale. In Europa, invece, esiste una normativa comunitaria che stabilisce un quadro unico e uniforme per tutti gli Stati membri: le legislazioni nazionali si basano sulla normativa europea adattandola alle specificità locali, ma rimanendo sempre allineate ai principi fondamentali dell'Unione Europea. Il fenomeno della globalizzazione e l'espansione del mercato dei servizi rendono indispensabile la creazione di un mercato unico tra i Paesi membri dell'Unione Europea, il quale richiede una riduzione della burocrazia e una maggiore armonizzazione delle normative economiche e politiche, evitando che le legislazioni nazionali divergano troppo tra loro. L'obiettivo è quindi quello di prevenire frammentazioni normative e promuovere una maggiore conformità alle leggi comunitarie, facilitando così l'integrazione economica e la competitività a livello globale.

3.2 Approccio economico europeo

In economia, il benessere totale o sociale è la somma del benessere dei consumatori e dei produttori. Tuttavia, all'origine delle leggi antitrust esisteva un dibattito su quanto peso dare al benessere dei produttori rispetto a quello dei consumatori. Il benessere dei consumatori dipende dalla soddisfazione che ottengono dai beni e servizi misurata attraverso il loro surplus, in relazione a qualità e prezzo. Questo surplus rappresenta la differenza tra quanto i consumatori sarebbero disposti a pagare per un bene e quanto effettivamente pagano. Il dibattito verteva sul fatto se, per il benessere economico complessivo della società, fosse più importante massimizzare solo il surplus dei consumatori o se considerare anche quello dei produttori.

Quando la Comunità europea è nata, il principale obbiettivo del diritto della concorrenza non era il benessere dei consumatori, bensì garantire la libera circolazione delle merci negli stati membri. Pertanto, per un lungo periodo ha primeggiato una visione più tradizionalista sulla protezione della concorrenza stessa, in cui l'obiettivo principale era garantire che i mercati rimanessero competitivi. Dunque, se un comportamento aziendale avesse danneggiato la concorrenza, la Corte avrebbe potuto considerarlo illegale senza la necessità di dimostrare che tale comportamento avesse causato un danno concreto o misurabile ai consumatori. In pratica, questo significa che la Corte non richiedeva necessariamente un'analisi economica dettagliata o modelli economici complessi per decidere se un'azione violasse il diritto antitrust. L'attenzione rimaneva sulla protezione del processo concorrenziale, non tanto sul dimostrare che i consumatori subivano un danno tangibile. Ciò riflette una visione più formale del diritto della concorrenza dove il mantenimento di quest'ultima è visto come un fine in sé, senza richiedere prove di conseguenze economiche specifiche. Solo dopo la fine degli anni 70' il diritto della concorrenza dell'Unione Europea ha iniziato ad avere un'impronta più economica, ponendo al primo posto la difesa di un mercato competitivo che consenta maggiore efficienza, con benefici sia per i consumatori che per le imprese. In un tale contesto concorrenziale si promuoveva un migliore sfruttamento delle risorse, in quanto le aziende si sentono in competizione tra di loro, e l'integrazione tra gli Stati membri, abbattendo le barriere commerciali tra i vari Paesi dell'Unione Europea. All'inizio degli anni 2000, l'aspetto economico del diritto della concorrenza è diventato predominante, con l'obiettivo principale di garantire il benessere dei consumatori. Ciò non significava eliminare solo le pratiche che danneggiavano direttamente questi ultimi, come l'aumento dei prezzi o la riduzione delle quantità di beni e servizi disponibili, ma significava anche intervenire contro le pratiche tra imprese che, pur non avendo un effetto immediato, potevano compromettere la concorrenza nel lungo periodo, alterando la struttura del mercato e riducendo i benefici che i consumatori avrebbero ottenuto da un mercato competitivo e dinamico. Sono stati integrati strumenti economici avanzati per valutare l'efficienza e misurare gli effetti concreti sul mercato delle pratiche aziendali. Anche se il benessere dei consumatori è una priorità centrale nella politica della concorrenza, la Corte Europea ha chiarito che essi non sono gli unici beneficiari di tale politica. Infatti, ogni attore del mercato, comprese le imprese, ha la responsabilità di adottare pratiche che non compromettano le dinamiche competitive in modo da assicurare che tutti i soggetti del mercato possano trarre vantaggio da un ambiente concorrenziale. Se le imprese operano in un mercato competitivo, esse stesse ne beneficeranno attraverso l'innovazione e l'efficienza, e

indirettamente anche i consumatori godranno dei vantaggi derivanti da una concorrenza sana e leale, come prezzi più bassi, maggiore qualità e varietà di prodotti.

Una delle principali pratiche concorrenziali sanzionate dalla politica della concorrenza europea è l'abuso di posizione dominante da parte delle imprese, vietato dall'articolo 82 del Trattato CE che forniva indicazioni su come affrontare tali casi basandosi su un'analisi economica più dettagliata e rigorosa. Questo articolo ha lo scopo di assicurare una concorrenza leale all'interno del mercato comune europeo, punendo le imprese che, abusando del loro potere, rischiano di danneggiare il mercato. Il "mercato" in questo contesto comprende sia le imprese che vi operano, poiché l'abuso di potere ne limita l'espansione e distorce la concorrenza, sia i consumatori, che, pur non subendo effetti immediati, potrebbero comunque vedersi danneggiati nel loro benessere a causa di una concorrenza indebolita. L'articolo non vieta l'acquisizione di una posizione dominante di per sé, se questa è ottenuta in modo leale grazie ai meriti dell'impresa. Tuttavia, l'impresa che raggiunge tale posizione ha la responsabilità di non abusarne per danneggiare la concorrenza nel mercato.

Successivamente, l'articolo 82 CE è stato sostituito dall'articolo 102 del Trattato sul Funzionamento dell'Unione Europea (TFUE), nel quale la Corte di Giustizia affronta il legame tra la libertà economica e la concorrenza. In particolare, afferma che la concorrenza è fondamentale per garantire la libertà economica, in quanto stimola l'attività economica e consente alle imprese la massima libertà di azione. L'idea di fondo è che un maggior livello di competizione conduce ad una maggiore varietà di scelta per i consumatori e, allo stesso tempo, favorisce la riduzione dei prezzi. Questo processo contribuisce a raggiungere l'efficienza allocativa, ovvero una distribuzione ottimale delle risorse, in cui il mercato risponde in modo più efficace alle esigenze e preferenze dei consumatori.

Esiste sempre un dibattito tra la scelta di una normativa diversificata e flessibile, in grado di adattarsi a casi specifici, oppure una legislazione univoca e omogenea che rendesse più forte e chiara l'applicazione delle leggi, specialmente nei mercati digitali, caratterizzati da dinamiche complesse e in rapida evoluzione. Ancora oggi non è chiaro quanto le istituzioni europee sostengano completamente l'"approccio più economico" nell'applicazione del diritto antitrust e quanto siano convinte che questo approccio vada bene per ogni singolo caso. Nel 2022, però, alcune sentenze del Tribunale (una corte inferiore alla Corte di giustizia) hanno fatto ampio uso di modelli economici per affrontare casi di abuso di posizione dominante, suggerendo una crescente apertura verso questo approccio.

3.3 Antitrust in Germania

Il diritto della concorrenza tedesco è poco noto al di fuori della Germania, probabilmente a causa della barriera linguistica e della percezione comune che lo considera una semplice applicazione delle norme comunitarie. Questo ha portato ad un maggior focus sullo studio delle leggi europee piuttosto che sulle specifiche attuazioni nazionali. Tuttavia, la normativa tedesca in materia di concorrenza presenta caratteristiche distintive e originali che la differenziano dalle altre giurisdizioni, rendendola meritevole di maggiore attenzione e studio. Il diritto antitrust tedesco ha indubbiamente influenzato lo sviluppo di quello comunitario, contribuendo a definirne molte linee guida. Tuttavia, mantiene un certo grado di autonomia, differenziandosi per alcuni aspetti e principi specifici, pur rimanendo strettamente connesso alle normative europee. Un autore americano, David J. Gerber, ha scritto un libro verso la fine degli anni Novanta circa il diritto della concorrenza tedesco che mostra una panoramica sull'Antitrust in Europa: si dimostra come il diritto della concorrenza europeo, in particolare quello tedesco, si sia sviluppato secondo correnti autonome e differenziate rispetto agli Stati Uniti, fungendo da modello di ispirazione per altri Paesi.

L'esperienza tedesca in Europa è la più datata, la legge Gesetz gegen Wettbewerbsbeschränkungen (GWB) entrata in vigore il 1° gennaio 1958 e influenzata dagli Stati Uniti dopo la Seconda Guerra Mondiale, veniva vista come strumento di liberazione politica ed economica per garantire la libertà e prevenire la concentrazione di potere economico che poteva nuovamente minacciare la democrazia dopo il periodo nazista. Tuttavia, anche se il diritto tedesco della concorrenza aveva un approccio economico derivato da quello anglosassone, si è sempre distinto da quest'ultimo per la sua visione più ampia. Sviluppatosi sulla base del pensiero ordoliberalista, tradizione giuridica e dottrinale molto ricca in materia di Antitrust che ha influenzato la comunità europea, il diritto della concorrenza tedesco credeva che per garantire l'equilibrio tra libertà economica e concorrenza è fondamentale che lo Stato intervenga attivamente attraverso l'implementazione di un solido quadro normativo mirato a creare e mantenere un ambiente di mercato regolato, dove la competizione tra imprese possa prosperare liberamente. Infatti, Il rischio di lasciare il mercato completamente autonomo senza un adeguato controllo statale è che possano emergere situazioni di **monopolio e oligopolio.** L'approccio tedesco al diritto della concorrenza, con il suo forte legame a principi di libertà individuale e giustizia economica, appariva diverso rispetto al modello americano che spesso pone maggiore enfasi sull'efficienza di mercato e sul libero gioco delle forze economiche. Nell'approccio tedesco, l'economia non si limita semplicemente a massimizzare l'efficienza, ma garantisce che ogni persona possa equamente partecipare al mercato. La divergenza tra gli Stati

Uniti e la Germania, pertanto, riguarda principalmente l'approccio alla politica antitrust: si discute se sia più vantaggioso adottare un intervento attivo dello Stato o affidarsi al libero mercato. Negli USA, si teme che un intervento eccessivo da parte dello Stato possa portare non solo a falsi negativi, ovvero situazioni in cui si omette di agire quando sarebbe necessario, ma anche a falsi positivi, ossia casi in cui aziende vengono ingiustamente accusate di comportamenti scorretti a causa di un'applicazione troppo severa delle norme. Di conseguenza, la legge statunitense tende a dare più fiducia alla capacità del mercato di autoregolarsi, suggerendo che senza l'intervento statale il mercato possa stabilire autonomamente un contesto competitivo. Tuttavia, è evidente che nella realtà i mercati non funzionano sempre in modo perfetto e che una politica regolatoria è spesso necessaria per garantire una concorrenza equa ed efficace.

Successivamente, la concezione ordoliberale in Germania della gestione dei mercati è stata parzialmente superata, poiché sono emerse alcune criticità legate alla sua visione rigida sul ruolo dello Stato e l'idea di eliminare completamente il potere di mercato. Col tempo, infatti, è stata riconosciuta la necessità di un approccio più bilanciato in cui, non solo lo Stato, ma anche il mercato stesso possa avere un ruolo attivo nella regolamentazione della concorrenza e nell'evoluzione economica. Comunque sia, fin dalla sua formazione il diritto della concorrenza tedesco si è contraddistinto da quello anglosassone ed europeo specialmente per l'elevata rigidità della componente normativa che induce ad una particolare attenzione alla chiarezza e completezza delle leggi. L'obiettivo è garantire che le regole siano precise e facilmente applicabili, permettendo una corretta e coerente osservanza. In effetti, le leggi tedesche nel campo di antitrust sono caratterizzate da un elevato dettaglio di descrizione e precisione: un particolare esempio è rappresentato dalle diverse sfaccettature della definizione di "posizione dominante" che viene data dal legislatore tedesco, la quale viene descritta nel dettaglio in base al contesto economico in cui ci si trova e viene continuamente aggiornata all'interno della legislazione antitrust. Nella riforma del GWB, entrata in vigore nel giugno 2017, vengono introdotte delle specifiche che possano aiutare le autorità antitrust a riconoscere ed eventualmente prevenire comportamenti anticoncorrenziali all'interno dei mercati digitali, e si discute su come interpretare le disposizioni legislative tedesche nel campo dell'antitrust. In particolare, si solleva la questione se sia più appropriato adottare criteri interpretativi standard, simili a quelli utilizzati in altre aree del diritto, o se sia necessario un approccio più autonomo e distintivo. Questo secondo approccio si caratterizza per un'interpretazione "più giuridica", che si differenzia dalle pratiche tradizionali e tiene conto delle specificità del diritto della concorrenza. Tuttavia, questa forte rigidità normativa era difficile da garantire e negli ultimi

anni è sempre più comune interpretare le norme nazionali in linea con il diritto antitrust europeo per garantire una coerenza tra le normative nazionali e quelle europee. Nonostante ciò, le legislazioni europee in materia di antitrust permettono un'interpretazione più autonoma e differente nella legislazione tedesca, che offre la possibilità di considerare ogni specifica situazione nel mercato. Inoltre, nel contesto dello sviluppo dei mercati digitali viene messa in evidenza la capacità innovativa e di visione prospettica dell'antitrust tedesco: ciò si riflette nelle proposte avanzate per la regolamentazione di tali mercati per la prevenzione di comportamenti scorretti da parte delle grandi aziende tecnologiche. L'approccio tedesco, in questo senso, ha mostrato un'attenzione proattiva nel cercare soluzioni adatte alle nuove sfide poste dall'economia digitale, anticipando problemi futuri e proponendo misure efficaci per garantire la concorrenza leale.

Un esempio significativo è il caso di Facebook, accusato nel 2019 dal Bundeskartellamt (FCO) di gestione scorretta dei dati personali degli utenti online. Questa accusa si inserisce in un programma più ampio avviato in Germania e successivamente sviluppato su più livelli, coinvolgendo ambiti accademici, politici, legislativi e amministrativi.

3.4 Caso Facebook in Germania

Nel marzo 2016 l'ufficio federale tedesco ha avviato un'indagine su Facebook. Già da tempo, la Germania osservava con sospetto le grandi aziende tecnologiche e aveva condotto numerose indagini sul loro trattamento dei dati degli utenti, soprattutto in relazione alle piattaforme multi-versanti che gestiscono diversi tipi di utenti e servizi. All'inizio la questione sembrava concentrarsi su una violazione della privacy: l'FCO (l'Autorità tedesca per la concorrenza) accusava Facebook di gestire i dati personali degli utenti in modo scorretto, non conforme alle regole stabilite dal GDPR, la normativa europea sulla protezione dei dati personali. L'FCO sosteneva che la piattaforma creava dei "super profili" attraverso la combinazione di informazioni provenienti da diverse fonti, sfruttando delle interfacce che vengono installate in altri servizi di sua proprietà. Di fatti, Facebook aveva recentemente acquisito Instagram e WhatsApp per poter accedere strategicamente ai dati personali degli utenti anche su queste piattaforme. In particolare, impone agli utenti, tra le clausole per l'uso del social network, di accettare la condivisione dei dati raccolti su WhatsApp e Instagram, integrando così le informazioni provenienti da tutte le piattaforme sotto il suo controllo. Facebook ha cercato di difendersi sostenendo che la creazione dei cosiddetti "super profili"

sia vantaggiosa per gli utenti. Secondo l'azienda, raccogliendo una vasta quantità di dati personali, inclusi quelli provenienti da fonti esterne alla piattaforma, è in grado di offrire pubblicità mirata e personalizzata che risponde meglio agli interessi degli utenti. Ha quindi affermato che questo sistema non costituisce uno sfruttamento, ma piuttosto un modello efficiente che porta benefici anche ai consumatori. La questione principale, secondo l'FCO, è che la creazione di questi "super profili" per migliorare l'efficienza della pubblicità personalizzata non giustifica l'uso di dati provenienti da fonti esterne a Facebook senza il pieno controllo degli utenti. In altre parole, l'efficienza vantata da Facebook non può prevalere sui diritti degli utenti, i quali, in questo contesto, non hanno la possibilità di decidere come e quando i propri dati personali vengono utilizzati. Di conseguenza, l'FCO ritiene che l'argomentazione di Facebook non possa essere accettata a causa del forte squilibrio tra profitti della piattaforma e i rischi ai quali sono esposti gli utenti. Questo tema, nonostante inizialmente sembra essere nato come un caso di violazione della privacy, ha anche implicazioni più ampie che coinvolgono il tema della concorrenza: Facebook, grazie a tali pratiche, avrebbe rafforzato la sua posizione dominante nel mercato digitale, distorcendo la concorrenza e limitando le opportunità per altri attori di competere equamente.

Nonostante la decisione dell'FCO, nel febbraio 2019, di ordinare a Facebook di porre fine a tale comportamento discriminatorio entro un arco temporale di 12 mesi, il sito web ha fatto ricorso contro l'Autorità alla Corte d'appello di Düsseldorf accusandola di non essere l'ente competente per occuparsi di un caso riguardante la privacy, poiché l'FCO è l'Autorità tedesca preposta alla regolazione della concorrenza. La Corte d'appello di Düsseldorf ha accolto la richiesta di Facebook, bloccando la decisione presa da parte dell'FCO, poiché ha sostenuto di non riconoscere l'accusa di posizione dominante nel mercato digitale. La Corte d'appello di Düsseldorf sostiene che un'accusa di abuso è valida solo quando sono chiaramente visibili effetti economici negativi, come un evidente aumento eccessivo dei prezzi. Tuttavia, nel caso presentato dall'FCO questo tipo di prova è assente, poiché il mercato digitale in questione ha un prezzo pari a zero, rendendo difficile utilizzare parametri quantitativi per dimostrare l'abuso. Inoltre, la Corte ritiene che Facebook non abbia adottato comportamenti diversi da quelli che avrebbe attuato in un mercato perfettamente concorrenziale. In altre parole, secondo la Corte, Facebook avrebbe imposto le stesse condizioni commerciali agli utenti anche se ci fosse stata una competizione equa. L'FCO, invece, sostiene che l'abuso di posizione dominante si verifica ogni volta che un'azienda impone determinate condizioni agli utenti senza offrire loro una reale possibilità di scelta, obbligandoli ad accettare tali condizioni per accedere ad un determinato servizio; questo per l'Autorità rappresenta un abuso

indipendentemente dal fatto che il mercato abbia un prezzo pari a zero. Dunque, l'FCO non si è arreso e ha presentato la questione davanti la Corte federale di giustizia tedesca (Bundesgerichtshof) con l'intento di ribaltare la sentenza della Corte d'Appello di Düsseldorf. Con successo, il 23 giugno 2020 il Senato dei cartelli della Corte federale ha ristabilito la validità della decisione dell'FCO rendendola nuovamente esecutiva, sebbene abbia leggermente modificato la motivazione del caso colmando alcune mancanze del suo ragionamento.

In primo luogo, la Corte federale ha spostato il focus dall'accusa di una semplice violazione della privacy a un concetto noto come "compulsory tying": Facebook costringe gli utenti ad accettare un servizio che non è strettamente necessario per loro. Gli utenti si iscrivono alla piattaforma principalmente per socializzare, scambiare messaggi e condividere contenuti con altri profili; tuttavia, oltre a fornire questa esperienza che gli utenti desiderano, Facebook impone anche un servizio altamente personalizzato, per il quale il "prezzo" richiesto è la cessione di una grande quantità di dati personali, anche se gli utenti non sanno esattamente come queste informazioni vengano gestite. In questo contesto, la Corte federale riconosce come Facebook, con la sua capacità di raccogliere enormi quantità di dati, detenga una posizione quasi monopolistica che gli permette di imporre condizioni che gli utenti non possono facilmente rifiutare, poiché non esiste una reale alternativa equivalente nel mercato. Si può concludere che questa pratica di "vendita abbinata" diventa illegale quando i due prodotti offerti insieme non sono intrinsecamente legati, ovvero uno non include necessariamente l'altro, ma l'azienda non permette ai consumatori di ottenere uno dei due senza accettare anche l'altro. Questa violazione risulta particolarmente evidente quando l'azienda occupa una posizione dominante sul mercato del prodotto principale, come nel caso di Facebook con il suo social network. Una prova chiave della natura anticoncorrenziale di questa pratica si verifica quando i consumatori per utilizzare solo il social network (prodotto principale) sono costretti a "pagare" con la cessione dei propri dati personali per avere pubblicità mirata (il prodotto secondario). Se l'azienda non offre l'opzione di poter usufruire solo del prodotto al quale i consumatori sono realmente interessati, e manca una proposta alternativa sul mercato, si rafforza ulteriormente l'idea di una restrizione della concorrenza.

In secondo luogo, la Corte federale tedesca ha fatto un'altra considerazione che va oltre la questione della protezione dei dati. Dal suo ragionamento viene fuori "l'idea del danno" riferita all'espansione imposta della prestazione: cioè, ci si riferisce al fatto che Facebook non si limita ad utilizzare solo dati degli utenti che si trovano su Facebook ma anche quelli al di fuori di Facebook, sfruttando servizi con i quali la piattaforma ha dei collegamenti.

Nonostante Facebook lo giustifica come una maggiore efficienza per l'utente, in realtà non è altro che una forma di sfruttamento a due livelli che, se da un lato personalizza ancor di più l'esperienza dell'utente, dall'altro lo sottopone a molti più rischi e lo costringe a "corrispondere" al servizio offerto con un costo molto più alto.

Per concludere, il Parlamento tedesco, dopo aver chiarito la questione del nesso di causalità, ha deciso di informare direttamente la Corte federale di giustizia tedesca circa le decisioni dell'FCO. La Corte d'appello di Düsseldorf viene esclusa dal processo di appello, riferito a questi casi nello specifico, per la battuta d'arresto che aveva fatto all'FCO e non aver considerato il caso nella sua complessità.

3.5 Causalità tra posizione dominante e abuso

Per dimostrare che Facebook abusava dei dati degli utenti per ottenere vantaggi commerciali, l'FCO doveva dimostrare che effettivamente il sito web godesse di una posizione dominante rispetto alle altre piattaforme nel mercato in questione. Secondo l'FCO, il servizio che Facebook offre all'utente è difficilmente replicabile dalle altre piattaforme, in quanto fornisce agli utenti un'esperienza quotidiana unica: la possibilità di connettersi con persone distanti, condividere messaggi, pensieri ed esperienze in una rete sociale globale. Valutando il servizio di altre piattaforme come YouTube, Twitter o Snapchat, l'Autorità tedesca mette in dubbio che possano costituire dei validi concorrenti, poiché non offrono la stessa ampiezza di servizi o lo stesso tipo di interazione sociale. Per questo motivo, l'FCO ha sostenuto che Facebook godeva di una posizione dominante nel mercato dei social network, la quale è difficile da dimostrare in un mercato in cui il prezzo per gli utenti è pari a zero. Infatti, per sostenere la sua tesi l'FCO ha adottato una metrica diversa per valutare la dominanza di Facebook: il numero di utenti attivi giornalieri registrati sulla piattaforma. Questo indicatore è stato considerato rilevante per dimostrare il controllo di Facebook sul mercato, poiché, nonostante la presenza di concorrenti, questi avevano avuto scarso successo nel mercato tedesco. Facebook ha cercato di difendersi affermando che il numero di utenti medi giornalieri fosse semplicemente uno "standard del settore". Tuttavia, la Corte federale e l'Autorità tedesca della concorrenza hanno rilevato che questi "standard" erano stati imposti proprio da Facebook, in quanto azienda leader del mercato: quando un'azienda dominante adotta certi comportamenti, è naturale che le imprese più piccole tendano a imitarli, facendoli diventare la norma. Di conseguenza, anche gli utenti possono perdere la consapevolezza delle alternative disponibili di un contesto concorrenziale, poiché condizioni di sfruttamento o limitazioni delle loro azioni possono apparire normali. Partendo da questo ragionamento, la Corte

federale e l'FCO hanno individuato un nesso causale tra il comportamento di Facebook e il consolidamento della sua posizione dominante, rilevando come tali pratiche abbiano influenzato negativamente il mercato e ridotto la concorrenza. La posizione dominante di un'azienda in un mercato comporta l'obbligo di rispettare in maniera particolare le norme sulla concorrenza, evitando comportamenti scorretti che possano distorcere il mercato a beneficio di pochi.

Una volta chiarito il potere di Facebook nel settore delle piattaforme digitali, il passo successivo era dimostrare l'abuso di tale potere per ottenere vantaggi commerciali. In effetti, l'FCO e la Corte federale tedesca hanno riscontrato che una chiara prova di abuso di potere consiste nel modo in cui Facebook sfrutta la propria capacità di raccogliere dati sugli utenti per offrire agli inserzionisti un servizio che le altre piattaforme non possono fornire, ovvero la possibilità di raggiungere un numero significativamente più elevato di potenziali clienti. Questo vantaggio deriva dalla vasta quantità di dati personali a cui Facebook ha accesso, mentre le altre piattaforme non possono vantare la stessa capacità. Tale fenomeno rappresenta un chiaro esempio di abuso di posizione dominante; infatti, le normative antitrust europee identificano l'abuso come l'imposizione di prezzi elevati ai consumatori o l'applicazione di condizioni commerciali scorrette. Nel caso di Facebook, l'azienda sfrutta condizioni commerciali particolari, come l'accesso a un'enorme mole di dati, che ostacola la concorrenza, poiché non tutte le aziende possono competere su un piano di parità. Pertanto, la misura dell'abuso in questo contesto non si basa tanto sulla perdita diretta di benessere del consumatore, considerando che si opera in un mercato a prezzo zero, quanto piuttosto sulle condizioni di disuguaglianza commerciale che vengono create. Nel contesto di Facebook, l'abuso che si accusa è quello di un utilizzo eccessivo dei dati rispetto a quanto sarebbe necessario in un contesto di concorrenza effettiva. Questo comportamento può essere paragonato all'abuso di prezzi eccessivi, ma invece di riferirsi al costo monetario, riguarda le condizioni imposte agli utenti per l'uso dei loro dati. In altre parole, è come se considerassimo i dati personali degli utenti una "moneta di scambio" per l'utilizzo della piattaforma. Una volta accertata la presenza di un abuso di posizione dominante, bisogna capire qual è la condizione esatta a partire dalla quale si può dichiarare un abuso di potere. Nel caso di prezzi eccessivi, trattandosi di grandezze quantitative, è più semplice per l'agenzia confrontare i prezzi considerati troppo alti con quelli di un mercato concorrenziale, valutandone il divario. Tuttavia, nel contesto di abuso di condizioni commerciali, come nel caso di Facebook, viene adottato un approccio diverso: piuttosto che basarsi su un confronto di prezzi, le Autorità utilizzano uno standard legale, facendo una particolare considerazione riguardo le grandi

aziende tecnologiche che non era mai stata formulata prima. In particolare, la Corte federale ha stabilito che quando un'azienda raggiunge una posizione dominante in un mercato ha una responsabilità particolare nel bilanciare i propri interessi economici con i rischi delle sue pratiche aziendali. In altre parole, deve evitare ogni comportamento scorretto che possa distorcere la concorrenza. Nel caso specifico di Facebook, la Corte ha ampliato questa responsabilità includendo non solo le regole della concorrenza ma anche il rispetto dei diritti costituzionali, solitamente associati agli enti pubblici, come il diritto alla privacy e all'autodeterminazione personale. Questo perché Facebook, oltre a essere una semplice azienda di social network, è diventato un pilastro fondamentale della comunicazione e dello scambio sociale nella società moderna. La sua piattaforma svolge un ruolo centrale nella vita quotidiana di milioni di persone, simile a quello di un'infrastruttura pubblica. Di conseguenza, la Corte ha affermato che Facebook deve essere trattato come una sorta di "infrastruttura sociale", quasi al pari di un ente pubblico, con l'obbligo di rispettare e proteggere i diritti fondamentali degli utenti; questo supera il semplice obbligo di rispettare le regole del mercato concorrenziale. In tale contesto, il GDPR diventa il punto di riferimento normativo per la tutela dei dati personali, e per tale motivo ogni violazione dei diritti sanciti dal GDPR da parte di Facebook e altre Big Tech non sarà solo un problema di concorrenza, ma costituirà anche un abuso di posizione dominante e una violazione dei diritti fondamentali degli utenti. Questo approccio sposta la questione oltre la semplice logica economica, riconoscendo l'importanza della protezione dei diritti personali nel mondo digitale.

3.6 Considerazioni finali sul caso

Il caso di Facebook è considerato uno dei più rilevanti in tema di regolamentazione europea sulla concorrenza poiché è uno dei pochi casi che mira ad attaccare il problema che sta alla base dell'era digitale con la nascita e lo sviluppo delle grandi piattaforme tecnologiche. Evidenzia per la prima volta il legame tra privacy e concorrenza, spingendo l'Antitrust a superare i tradizionali confini economici, che solitamente si concentrano esclusivamente sugli aspetti di concorrenza. Con tale approccio è stata particolarmente attirata l'attenzione di molti studiosi americani poiché evidenzia chiaramente le differenze tra il diritto antitrust statunitense e quello in Europa, in particolare in Germania. Nel diritto della concorrenza anglosassone gli economisti si basano principalmente sul benessere dei consumatori come metro di paragone per misurare l'efficienza; pertanto, quando nel 2016 l'Antitrust tedesca ha avviato il caso contro Facebook, risultava difficile comprendere quale fosse il reale danno economico, a livello quantitativo, che l'agenzia voleva provare. Invece, la procedura giuridica adottato dall'Autorità

antitrust tedesca, nota come "bilanciamento normativamente orientato degli interessi", non si limitata a valutare esclusivamente gli aspetti economici del caso, ma cerca di mettere in equilibrio, da un lato, gli interessi dell'azienda e, dall'altro, i diritti degli utenti, in particolare per quanto riguarda la protezione dei loro dati personali. Questa decisione segna una svolta importante, poiché per la prima volta in un caso di abuso di posizione dominante viene considerata la violazione della concorrenza insieme alla normativa sulla privacy. Questo tipo di valutazione normativa è una caratteristica distintiva del diritto antitrust tedesco, che cerca spesso di integrare principi giuridici più ampi e di dare un peso significativo a norme extra-economiche, come in questo caso la legislazione sulla protezione dei dati personali. Questo approccio ha reso più facile per l'Autorità dimostrare che la condotta di Facebook, il quale imponeva agli utenti di accettare la condivisione dei dati personali raccolti su altre piattaforme (come WhatsApp e Instagram), era effettivamente abusiva, con l'integrazione di normative relative alla privacy.

Altra considerazione importante sul caso è che la scelta della Corte federale tedesca di basare l'analisi sull'articolo 19 del GWB (Gesetz gegen Wettbewerbsbeschränkungen), che corrisponde all'articolo 102 del Trattato sul Funzionamento dell'Unione Europea (TFUE), offre una prova significativa del fatto che il diritto tedesco ha una giurisprudenza e una dottrina più consolidate e favorevoli rispetto al diritto europeo. Infatti, l'articolo 19 del GWB proibisce l'abuso di posizione dominante nel mercato tedesco, analogamente a quanto previsto dall'articolo 102 TFUE a livello comunitario. Inoltre, proprio l'articolo 19 del GWB è stato modificato, sotto proposta del governo tedesco, pochi mesi dopo il rifiuto da parte della Corte d'appello riguardo la decisione dell'FCO. Infatti, precedentemente a tale modifica, per dimostrare l'abuso di sfruttamento era necessario provare un legame di causalità diretto tra la condotta dell'impresa dominante e il danno subito dalla controparte. Con la riforma, invece, il legislatore ha semplificato questa regola permettendo di considerare come abuso di sfruttamento anche situazioni in cui la parte dominante impone condizioni sfavorevoli e ingiustificate, senza dover dimostrare un nesso causale stretto.

Sebbene le accuse avanzate dall'Antitrust tedesca contro Facebook siano state incisive, non sono ancora state prese misure definitive per tutelare efficacemente la posizione degli utenti online. Questo ritardo è dovuto alla complessità delle procedure legali e non è necessariamente una colpa delle agenzie, poiché i processi legali richiedono tempo per produrre soluzioni innovative e ben ponderate, e affrettare le decisioni potrebbe portare a risultati meno efficaci e rischiosi. Comunque sia, ogni fase del processo è stata cruciale per

arricchire il dibattito e arrivare a una decisione più solida; le critiche mosse dalla Corte d'Appello nei confronti delle decisioni iniziali dell'FCO hanno infatti evidenziato alcune lacune nella valutazione dell'Antitrust. Queste osservazioni sono servite a migliorare la riflessione giuridica, portando la Corte di Giustizia dell'Unione Europea ad intervenire e colmare quei vuoti con una decisione finale di grande portata. Questo processo dimostra che, pur richiedendo tempo, la riflessione graduale e il confronto tra diverse fasi procedurali sono essenziali per giungere a esiti più significativi, in linea con i diritti fondamentali degli utenti e le nuove sfide imposte dalle piattaforme tecnologiche.

Dunque, sebbene il caso di Facebook non abbia risolto completamente il problema dell'abuso di posizione dominante, né introdotto soluzioni immediate e impattanti, ha comunque avuto un ruolo fondamentale nell'evidenziare un aspetto finora trascurato: il legame tra la protezione della privacy e il diritto della concorrenza. Prima di questo caso, tali ambiti erano stati considerati in modo separato, ma attraverso le accuse mosse contro la piattaforma sono emerse nuove criticità e lacune nelle normative esistenti. Grazie a questo caso, si è innescato un dibattito che ha portato a revisioni e aggiornamenti delle normative, culminando nella creazione di nuovi strumenti regolatori: il Parlamento europeo ha emanato il Digital Markets Act (DMA), un nuovo quadro normativo proprio per affrontare il potere delle grandi piattaforme digitali e introdurre regole più rigide su come queste aziende devono gestire dati e concorrenza. In particolare, l'art. 5 (2) (b) della normativa dispone che il gatekeeper non deve:

"combinare i dati personali del servizio di piattaforma principale pertinente con i dati personali di qualsiasi altro servizio di piattaforma principale o di qualsiasi altro servizio fornito dal gatekeeper o con i dati personali di servizi di terze parti (...) a meno che all'utente finale non sia stata presentata la scelta specifica e non abbia prestato il consenso ai sensi del [GDPR]."

In definitiva, pur essendo considerato da alcuni come un fallimento per quanto riguarda l'impatto diretto sui mercati, il caso ha svolto una funzione cruciale nel riorientare l'attenzione politica e legale. Ha spinto le autorità a rivalutare e migliorare la progettazione istituzionale del diritto economico, contribuendo così a un'evoluzione normativa che avrà effetti nel lungo termine sul modo in cui i colossi tecnologici operano.

3.7 I confini del diritto Antitrust

Uno dei principali dibattiti sul caso Facebook riguarda l'FCO, ovvero se l'autorità Antitrust tedesca fosse effettivamente l'organo competente per affrontare tale caso. Questo perché in

principio il problema appariva principalmente come una questione di violazione della privacy, senza un legame chiaro e diretto con la concorrenza. Infatti, molti critici ritenevano che il diritto antitrust non fosse lo strumento adeguato ad affrontare problemi sociali complessi come la privacy o la disuguaglianza. Si pensava che le competenze dell'antitrust fossero più limitate, ovvero incentrate sulla tutela della concorrenza e sul mantenimento dei mercati competitivi, senza la capacità di intervenire in questioni che esulano dalla sua sfera, come la protezione dei dati personali o altre sfide sociali. Ecco il motivo del rifiuto da parte della Corte d'Appello tedesca della decisione dell'FCO, perché era riluttante nel credere che la cattiva gestione della privacy da parte di Facebook avesse realmente danneggiato la concorrenza, cioè non riusciva a comprendere gli effetti economici negativi che derivavano dall'abuso di potere da parte di Facebook. Da qui nasce la critica fatta all'FCO: si credeva che occupandosi di questioni di privacy avesse oltrepassato i limiti dell'area di sua competenza, per sfociare in un ambito che dovrebbe essere gestito dalle agenzie dedicate alla protezione della privacy e non dalle autorità antitrust.

Il dibattito iniziale sul caso Facebook ruotava attorno alla convinzione diffusa che i campi della concorrenza e della privacy fossero completamente separati. Tuttavia, il caso ha portato a una riflessione profonda su questa presunta disconnessione, evidenziando come, con l'espansione dei mercati digitali, i dati personali siano diventati uno dei principali asset competitivi. È emerso chiaramente che il controllo e l'accesso a grandi quantità di dati conferisce un vantaggio competitivo significativo, rendendo i dati una risorsa cruciale per la competizione sul mercato. Pertanto, non è improprio che le autorità antitrust intervengano su questioni legate alla privacy, poiché la gestione dei dati personali incide direttamente sulla dinamica competitiva.

La privacy, in questo contesto, non è solo una questione sociale o etica, ma diventa un elemento essenziale nella regolamentazione della concorrenza. Di conseguenza, l'Antitrust ha il compito legittimo di occuparsi dei problemi relativi alla privacy, poiché questi riguardano una risorsa fondamentale nella competizione tra aziende. È importante riconoscere che, sebbene la concorrenza non sia la soluzione definitiva per risolvere tutte le problematiche sociali, come la privacy o la disuguaglianza, essa contribuisce comunque a evitare che questi problemi si aggravino. Nel caso di Facebook, ad esempio, è stato confermato che la piattaforma detiene una posizione dominante nel mercato dei social network, con poche alternative per gli utenti che offrano servizi simili. Si è dimostrato che, se ne avessero la possibilità, molti utenti sceglierebbero piattaforme che richiedano meno dati personali; tuttavia, l'assenza di una concorrenza reale limita questa possibilità, poiché non ci sono altre

piattaforme che rispondano a tale richiesta. In altre parole, la mancanza di concorrenza non solo riduce la scelta per gli utenti, ma impedisce anche l'emergere di offerte che soddisfino esigenze diverse, come una maggiore protezione della privacy. Questo dimostra che la concorrenza non influisce solo su aspetti economici, ma tocca anche ambiti più ampi, come la tutela dei diritti dei consumatori. In un contesto privo di concorrenza, altri valori fondamentali, come la privacy e la trasparenza, rischiano di deteriorarsi. Dunque, un mercato concorrenziale non è solo un fattore economico, ma una condizione che può influenzare positivamente molte dimensioni sociali.

CONCLUSIONE

Il presente elaborato mette in luce il conflitto emerso nell'era digitale, caratterizzata dalla crescente valorizzazione dei dati, tra due esigenze percepite come contrastanti: la protezione della privacy e la promozione della crescita economica attraverso la concorrenza.
Inizialmente, vi era una forte resistenza a riconoscere un collegamento tra questi due campi; attraverso il caso di Facebook in Germania, invece, vengono messe in risalto le diverse sfaccettature e strade che il diritto della concorrenza può intraprendere, estendendosi a contesti molto più ampi e complessi dei semplici strumenti economici quantitativi. Con il tempo, gli strumenti, le procedure e gli obbiettivi del diritto Antitrust si sono evoluti, mirando non più al semplice benessere economico attraverso la salvaguardia del contesto concorrenziale, ma occupandosi di tanti altri aspetti come il benessere dei consumatori, la struttura del mercato, l'equità e l'integrazione europea.

Il conflitto tra privacy e concorrenza nasce dal fatto che le normative sulla protezione dei dati personali sono talvolta considerate come un freno alla crescita economica, in quanto vengono ritenute in grado di limitare la capacità delle imprese di utilizzare i dati degli utenti per finalità che potrebbero sembrare vantaggiose per tutti gli attori del mercato. Secondo questa visione, tali leggi riducono la competitività impedendo alle aziende di sfruttare appieno il potenziale dei dati per migliorare i propri servizi, ottimizzare l'offerta e, di conseguenza, favorire l'innovazione. Si crea così una tensione tra la necessità di proteggere la privacy degli individui e quella di consentire alle imprese di operare in un ambiente economico dinamico e competitivo, in cui i dati diventano un elemento cruciale per la crescita e la competitività. Molto spesso la letteratura economica non riesce a contestualizzare correttamente gli effetti delle normative, concentrandosi su aspetti limitati nel breve periodo e trascurando un'analisi più completa a lungo termine. Infatti, uno degli errori più comuni è considerare solo l'impatto

diretto delle normative sulle aziende, ad esempio sui profitti pubblicitari, senza valutare gli effetti sul benessere dei consumatori e sull'intera società.

Nel dibattito sulla privacy, il problema principale è la difficoltà nel comprendere e quantificare con precisione i danni derivanti dalla violazione dei dati personali. Non è ancora chiaro come il valore generato dai dati venga distribuito tra i diversi soggetti coinvolti, e questa mancanza di conoscenza, unita all'incertezza sui reali danni alla privacy, fa sembrare che la protezione della privacy e il valore dei dati siano in conflitto. Tuttavia, si solleva la questione che tale conflitto sia stato eccessivamente semplificato, riducendolo ad una contrapposizione netta. Si mette in dubbio che debba esserci un vincitore tra i due, senza considerare la possibilità che privacy e utilizzo economico dei dati possano coesistere. L'esperienza dimostra che le normative sulla privacy, seppure implementate per tutelare gli utenti, non hanno fermato la crescita economica dei dati, né ridotto il valore economico generato. Anzi, le preoccupazioni dei consumatori riguardo alla protezione della loro privacy persistono, segnalando che il tema resta irrisolto e complesso. Questa confusione ha portato alla percezione diffusa che le attuali siano costose e inefficaci nel proteggere la privacy degli utenti, apparendo addirittura inutili perché non risolvono il problema di fondo. Le normative, dunque, vengono spesso viste come degli ostacoli al valore economico dei dati, ma questo fraintendimento nasce da una visione poco chiara del concetto di privacy e della sua complessità. Comunque sia, la privacy è un qualcosa che, nonostante sia stato sempre presente nell'essere umano, si è evoluto con l'evolversi della società e la nascita di nuove esigenze. Gli economisti continuano ad analizzare il problema sempre dallo stesso punto di vista: non è corretto mantenere rigorosi metodi di ricerca quando viene ampliato l'orizzonte di indagine. Affrontare la privacy esclusivamente da un punto di vista economico può chiarire alcuni aspetti della questione, ma comporta anche il rischio di trascurare importanti dimensioni sociali e psicologiche.

In conclusione, le considerazioni economiche hanno fortemente influenzato le regolamentazioni sulla privacy, riducendo il concetto al punto da trascurare quasi del tutto aspetti cruciali come la libertà e i diritti fondamentali. Di conseguenza, tali normative risultano frammentate e limitate, concentrandosi principalmente su meccanismi di notifica e consenso, senza fornire strumenti reali e concreti per proteggere la privacy come un autentico diritto fondamentale. Si ha la necessità di sviluppare strumenti più efficienti nel proteggere i dati degli utenti online, ma per far ciò bisogna rispondere alle domande su come venga distribuito il valore dei dati e chi ne trae realmente vantaggio.