Abstract

La crescita di Internet e la pervasività delle tecnologie dell'informazione e della comunicazione (TIC) hanno portato a un cambiamento radicale nella nostra società, un profondo impatto economico, commerciale e sociale sulle nostre vite. Ad oggi, la maggior parte della nostra vita si svolge online dove gli algoritmi modellano e guidano il nostro comportamento e governano le nostre società.

Uno degli svantaggi di questo cambiamento è un aumento del rischio per gli utenti di Internet per la *privacy* delle loro informazioni personali. In effetti, un'enorme quantità di dati viene generata e diffusa da persone a ritmi elevati, spesso senza sapere chi sta registrando e cosa su di loro. La navigazione online, le operazioni bancarie, gli acquisti, le interazioni con i social network e qualsiasi tipo di collaborazione e comunicazione economica, sociale e personale online potrebbero minare la privacy degli individui a causa di una varietà di fattori che includono non solo lo spaventoso aumento della fuga di informazioni. In effetti, informazioni private specifiche possono anche essere dedotte/estratte tramite euristiche computazionali applicate ai dati (apparentemente non correlati a tali informazioni) che gli utenti divulgano volontariamente su Internet.

In particolare, tali falle di privacy possono essere causate sia da *(a)* applicazioni o software che gli utenti utilizzano intenzionalmente inconsapevoli dei rischi correlati, e *(b)* da pratiche dannose (illegali o sleali) perpetrate furtivamente da "avversari". Pertanto, la protezione dei dati privati, dei dispositivi e della privacy degli utenti nella società digitale è diventata una delle massime preoc-

cupazioni per individui, organizzazioni, governi nazionali e ricercatori.

Data la complessità, l'imperscrutabilità del software con cui gli utenti interagiscono e il numero di potenziali attacchi e pratiche sleali in cui possono incorrere, diventa sempre più chiara la necessità di salvaguardie tecnologiche che supportino gli utenti nel rilevamento e nel contrastare tali minacce. A questo proposito, il *machine learning* (ML), con la sua capacità di riconoscimento dei pattern, sembra essere un prezioso alleato in grado di difendere non solo la privacy degli utenti ma anche altri diritti minacciati nella società digitale.

In questa tesi, *ci si concentra* su tutele intelligenti per la *privacy* degli utenti nella società digitale. L' *obiettivo* è esplorare, sia a livello teorico che sperimentale, come gli approcci ML possono essere adattati a supportare la protezione della privacy e dei diritti correlati. La ricerca attinge anche ai più recenti sviluppi nei settori della computational law e della tecno-regolazione, due paradigmi di ricerca emergenti, su scala planetaria, ai confini tra informatica e diritto.

La tesi è strutturata come segue. Si descrive prima il quadro teorico e metodologico di questo lavoro, attraverso una revisione sistematica della letteratura che inquadra l'uso del ML per proteggere la privacy degli utenti. In questo modo, si riconducono gli approcci e le soluzioni esistenti per la protezione della privacy a due categorie fondamentali: *enforcement* (ovvero, soluzioni che *impongono vincoli* e ostacolano le violazioni delle norme) e *nudge* (cioè soluzioni che informano gli utenti e aumentano la loro *awareness* per promuovere comportamenti orientati alla privacy). Si fornisce una tassonomia completa delle principali aree, minacce, metodi ML, tipo di protezione fornita analizzando 143 studi pubblicati da Gennaio 2017 a Ottobre 2020.

Quindi, si presenta una serie di attività di ricerca che esplorano l'applicabilità di approcci basati sul ML ai problemi che sorgono negli scenari sopra descritti. Le attività presentate possono essere idealmente suddivise in due parti incentrate rispettivamente sulla tutela della privacy e altri diritti correlati. Più in dettaglio, la

prima parte comprende due progetti che affrontano la tutela della privacy in senso stretto.

a) ML per privacy enforcement
Ci si occupa dell'annoso problema del monitoraggio di terze parti sul Web in cui i dati privati degli utenti vengono rubati ingiustamente per attività di marketing e dannose, come lo stalking online. Si sperimenta l'uso del ML per distinguere tra tracker e risorse funzionali sul Web, scoprendo che queste tecniche possono essere utilizzate con un alto tasso di riconoscimento di tali minacce. L'approccio risultante basato sul ML è stato implementato in *GuardOne*, uno strumento per proteggere gli utenti dal monitoraggio di terze parti, che fornisce soluzioni *enforcement* per bloccare i tracker. *GuardOne* è stato valutato nel mondo reale rispetto a diverse soluzioni commerciali simili per la protezione della privacy. Le caratteristiche principali di *GuardOne* sono: *(i)* un meccanismo ibrido basato su ML e blacklist, *(ii)* personalizzazione delle blacklist basata sulle abitudini di navigazione dell'utente, *(iii)* a implementazione molto leggera che non influisce sulle prestazioni dei dispositivi degli utenti rispetto alle soluzioni commerciali, *(iv)* un'elevata efficacia nel rilevare e bloccare i tracker di terze parti, meglio della maggioranza delle soluzioni commerciali.

b) ML per privacy awareness
Si affronta del tema della diffusione inconsapevole e/o incontrollata di dati personali e privati, in formato testuale, su Internet. Si sperimenta l'utilizzo del ML e tecniche avanzate di elaborazione del linguaggio per supportare sia la classificazione dell'argomento del testo (tra i più sensibili, es. politica e salute) sia la sensibilità del contenuto in base a tale argomento, trovando che le prestazioni della soluzione proposta in un ambiente simulato sono paragonabili alle soluzioni disponibili in letteratura. Inoltre, si sperimenta come le soluzioni ML progettate possano essere adatte per apprendere le attitudini personali dell'utente nei confronti della privacy. Si sono quindi incorporati tali approcci di ML in *Knoxly*, uno strumento per proteggere gli utenti dalla diffusione di informazioni personali e/o private online, che si basa su soluzioni di *nudge*. In

particolare, *Knoxly* mira a sensibilizzare e promuovere comportamenti orientati alla privacy mediante avvisi/avvertimenti. Le caratteristiche principali di *Knoxly* sono: *(i)* un modulo Keyword per rilevare parole sensibili comuni e Personally Identifiable Information, *(ii)* un modulo Topic per distinguere l'argomento del testo, *(iii)* un modulo Sensibilità per "misurare" la sensibilità del contenuto del testo, *(iv)* un modulo personalizzato che consente all'utente di personalizzare gli avvisi visualizzati, *(v)* un'interfaccia utente intuitiva alimentata dalle tecniche di Visual Analytics, *(vi)* un'implementazione leggera che non influisce sulle prestazioni dei dispositivi degli utenti.

La seconda parte comprende altri due progetti che sfruttano soluzioni e approcci metodologici e tecnologici identificati nella fase di ricerca precedente e ampliano l'ambito della ricerca applicando tali intuizioni ad altri diritti, legati alla privacy e di grande rilievo nella società digitale, ovvero protezione dei minori e dei consumatori.

a) ML per child protection
Si affronta la sfida di fornire tutele online per una categoria specifica di utenti, ovvero i bambini, che, secondo il Regolamento generale sulla protezione dei dati (GDPR) e l'UNICEF, necessitano di misure di sicurezza *ad-hoc*. All'interno di questo progetto di ricerca, denominato *AI4Children*, si sperimentano diversi approcci basati sul ML per l'identificazione degli utenti, i quali possono essere visti come una base per sostenere gli standard legali per la protezione dei minori online. Una volta individuato l'utente, infatti, è possibile attivare le tutele specifiche. Più in dettaglio, l'approccio concepito (basato su tecniche di integrazione dei dati) è in grado di riconoscere l'età di un utente in base ai gesti tattili che esegue su un dispositivo mobile con un'elevata precisione. Le caratteristiche principali di *AI4Children* sono: *(i)* distinguere tra adulti e minorenni in base a gesti tattili comunemente eseguiti, *(ii)* utilizzare un piccolo insieme di caratteristiche e gesti tattili per una classificazione accurata, *(iii)* robustezza nella classificazione su diversi dispositivi.

b) ML for consumer protection
Si affronta la questione delle clausole sleali nei Termini di servizio online (ToS), ovvero clausole che minacciano direttamente gli interessi concreti degli utenti, ad esempio regolando il modo in cui i dati saranno gestiti e le responsabilità su di essi. Sperimentiamo l'uso di ML e di tecniche avanzate di elaborazione del linguaggio per supportare la classificazione delle categorie di clausole ToS e il livello di equità, scoprendo che la soluzione proposta supera soluzioni presenti in letteratura e può essere utilizzata per misurare la mancanza di chiarezza nei ToS. L'approccio concepito è stato incorporato in *ToSware*, uno strumento per aumentare la *consapevolezza* dei consumatori contro le pratiche sleali nei ToS online. Le caratteristiche principali di *ToSware* sono: *(i)* avere un meccanismo per misurare l'ingiustizia dei ToS online, *(ii)* rendere i ToS più facili da leggere grazie a diverse tecniche di visualizzazione e metafore visuali valutate da utenti reali, *(iii)* un'implementazione leggera che non influisce sulle prestazioni dei dispositivi degli utenti.

La tesi si conclude con considerazioni sulle sfide per la ricerca nel ML in queste aree specifiche e le prospettive future dispiegate dalla computational law e dalla tecno-regolazione.