



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Informatica

Corso di Laurea Triennale in Informatica

TESI DI LAUREA

Un Approccio Rule-based per l'Identificazione del Dark Pattern Trick Question

PRIMO RELATORE

Prof. Fabio Palomba

SECONDO RELATORE

Dott.ssa Giulia Sellitto

Università degli Studi di Salerno

CANDIDATO

Antonio Scognamiglio

Matricola: 0512109178

Anno Accademico 2021-2022

L'unico modo di fare un gran bel lavoro è amare quello che fate.

Steve Jobs

Abstract

Al giorno d'oggi le interfacce utente sono prodotti di lavoro interdisciplinare, in cui sono coinvolti non solo programmatore ma anche psicologi, progettisti di grafica, ingegneri dell'er-
gonomia umana, antropologi e sociologi. Per questo motivo gli elementi presenti nelle UI,
le loro posizioni, colori e forme non sono scelti casualmente, ma sono frutto dell'accurato
lavoro di un team composto da professionisti in diversi ambiti.

Molto spesso non ce ne accorgiamo, ma durante l'utilizzo di siti web o app, le nostre scelte
e azioni sono influenzate dai *dark pattern*, ovvero dei modelli riutilizzabili di design, che in
qualche modo fanno sì che l'utente compia delle scelte che normalmente non avrebbe fatto.
Con questo studio si vuole realizzare un algoritmo che supporti l'utente nell'individuazione
del dark pattern *trick question*, da integrare successivamente con un tool che con altri algoritmi
permette di identificare un insieme di diversi dark pattern.

Indice

Indice	ii
Elenco delle figure	iv
Elenco delle tabelle	vi
1 Introduzione	1
1.1 Contesto applicativo	2
1.2 Motivazioni e Obiettivi	4
1.3 Risultati ottenuti	5
1.4 Struttura della tesi	5
2 Stato dell'arte	7
2.1 Tassonomia	7
2.1.1 Nagging	8
2.1.2 Obstruction	10
2.1.3 Sneaking	12
2.1.4 Interface Interferences	16
2.1.5 Forced Action	19
2.2 Impatto dei Dark Pattern	22
2.3 Implicazioni sulla privacy	23
2.4 Identificazione dark pattern	24

3 Un approccio rule-based per l'identificazione del dark pattern Trick Question	26
3.1 Analisi del dataset	27
3.2 Implementazione dell'algoritmo	30
4 Valutazione preliminare dell'approccio	33
4.1 Metodologia	33
4.2 Risultati	34
5 Conclusioni e sviluppi futuri	36
Ringraziamenti	37
Bibliografia	39

Elenco delle figure

1.1	Esempio di utilizzo di un dark pattern su <i>postoffice.co.uk</i>	2
1.2	Esempi di utilizzo di dark pattern in siti e app	3
1.3	Intervento di Katerine Zou il 16 Marzo 2022 alla IMCO	4
2.1	Esempio del dark pattern <i>nagging</i> nell'app <i>Youtube</i>	9
2.2	Esempio del dark pattern <i>roach motel</i> sul sito <i>Spotify</i>	10
2.3	Esempio del dark pattern <i>price comparison prevention</i> sul sito <i>AliExpress</i>	11
2.4	Esempio del dark pattern <i>intermediate currency</i> nel gioco <i>Heroes and Generals</i> .	12
2.5	Esempio del dark pattern <i>forced continuity</i> nel sito <i>skillshare.com</i>	13
2.6	Esempio del dark pattern <i>hidden costs</i> nel sito <i>broadway.com</i>	14
2.7	Esempio del dark pattern <i>sneak into basket</i> nel sito <i>godaddy.com</i>	15
2.8	Esempio del dark pattern <i>bait and switch</i> nel sito <i>reddit.com</i>	16
2.9	Esempio del dark pattern <i>hidden information</i> nel sito <i>rac.co.uk</i>	17
2.10	Esempio del dark pattern <i>toying with emotion</i> nel sito <i>wish.com</i>	18
2.11	Esempio del dark pattern <i>false hierarchy</i> nel sito <i>lastminute.com</i>	19
2.12	Esempio del dark pattern <i>trick questions</i> nel sito <i>sky.com</i>	19
2.13	Esempio del dark pattern <i>social pyramid</i> nell'app <i>Farmville</i>	20
2.14	Esempio del dark pattern <i>privacy zuckering</i> nell'app <i>Messenger</i>	21
2.15	Esempio del dark pattern <i>gamification</i> nell'app <i>Candy Crush Saga</i>	22
3.1	Illustrazione metodologia utilizzata	27
3.2	Corrispondenza tra DOM e pagina di registrazione del sito www.very.co.uk .	31
3.3	Risultati generati dall'algoritmo mostrati nel terminale	32

- 4.1 *Trick question* implementato attualmente su www.very.co.uk 35

Elenco delle tabelle

2.1 Tabella di riepilogo riguardo l'identificazione dei dark pattern	25
3.1 Frasi etichettate come <i>trick question</i> nel dataset rilasciato da Mathur <i>et al.</i> [2019]	29

CAPITOLO 1

Introduzione

Attualmente tutte le aziende che si interfacciano ai clienti tramite siti web o app, lavorano sulla UX (User Experience) dei loro prodotti. La UX è una di quelle discipline che si applica alla progettazione di interfacce utente che mira a rendere più intuitiva e scorrevole l'esperienza dell'utilizzatore con esse.

Il motivo dietro questa scelta è rendere più efficace l'interazione tra l'utente e il prodotto, in modo da, tra le altre cose, aumentare la fidelizzazione dei clienti, che continueranno ad utilizzare il prodotto in virtù della piacevole esperienza che hanno quando si interfacciano con esso; questo conferisce prestigio e credibilità al brand e di conseguenza i guadagni dell'azienda aumentano.

In generale, lo scopo della progettazione UX è aumentare l'usabilità di un interfaccia; lo standard ISO definisce l'usabilità come "*l'efficacia, l'efficienza e la soddisfazione con le quali determinati utenti raggiungono determinati obiettivi in determinati contesti*".

Per massimizzare l'usabilità si può usare l'*HCI* (Human-Computer Interaction), ovvero lo studio dell'interazione tra gli esseri umani e i computer per la realizzazione di sistemi interattivi che siano usabili, affidabili e che supportino e facilitino le attività umane; questa disciplina fornisce una serie di strumenti da applicare durante la progettazione di un sistema, come *principi di usabilità, euristiche, linee guida e design pattern*. Molto spesso, però, per una questione di profitti, si tende ad ignorare questo tipo di approccio, per favorirne uno che possa fornire i suoi frutti in tempo molto breve, e cioè, l'introduzione di dark pattern; così facendo, si tenta di ingannare l'utente in modo da fargli condividere più dati di quelli che



Figura 1.1: Esempio di utilizzo di un dark pattern su *postoffice.co.uk*

vorrebbe o far si che esso spenda soldi anche se non è necessario.

1.1 Contesto applicativo

Con il termine *dark pattern* si fa riferimento a quelle interfacce utente o percorsi di interazione con un servizio, appositamente progettati per guidare l'utente finale verso comportamenti da questo non realmente desiderati.

I "comportamenti" non realmente desiderati dall'utente molto spesso sono quelli che poi portano un beneficio all'azienda che fornisce il prodotto, o in termini economici, o in termini di fidelizzazione.

Un esempio dell'implementazione di un dark pattern è mostrato in Figura 1.1; nello specifico si tratta dell'azienda *Post Office* (istituto di credito inglese) che per far scegliere all'utente se ricevere notifiche o meno utilizza delle *checkbox* rotonde in modo che sembrino dei *radio button*, lasciando intendere che le scelte siano mutuamente esclusive.

Un'altra azienda che ha utilizzato un dark pattern è *Optin Monster* come si può notare in Figura 1.2a: in questo caso si cerca di indurre un senso di vergogna nell'utente, che per rifiutare l'offerta deve cliccare sul bottone il cui testo è "*No grazie mi sta bene perdere clienti*". Quest'ultimo esempio è incluso nella classe dei dark pattern che vengono implementati tramite **manipolazioni** di frasi, parole, o testo in generale.

Molto spesso i dark pattern prendono di mira anche la privacy dell'utente, come nell'esempio in Figura 1.2b, dove Whatsapp implementa un flow di UX per informare l'utente dell'aggiornamento dei termini sulla privacy alquanto confuso: nella schermata 1 (in figura a sinistra), Facebook, società che controlla Whatsapp, non viene menzionata in nessun modo; solo cliccando su "read more" per visualizzare più dettagli, si scopre che l'utente, utilizzando l'app, deve fornire il consenso anche a Facebook di poter usare i dati raccolti da Whatsapp su di esso a fini promozionali.

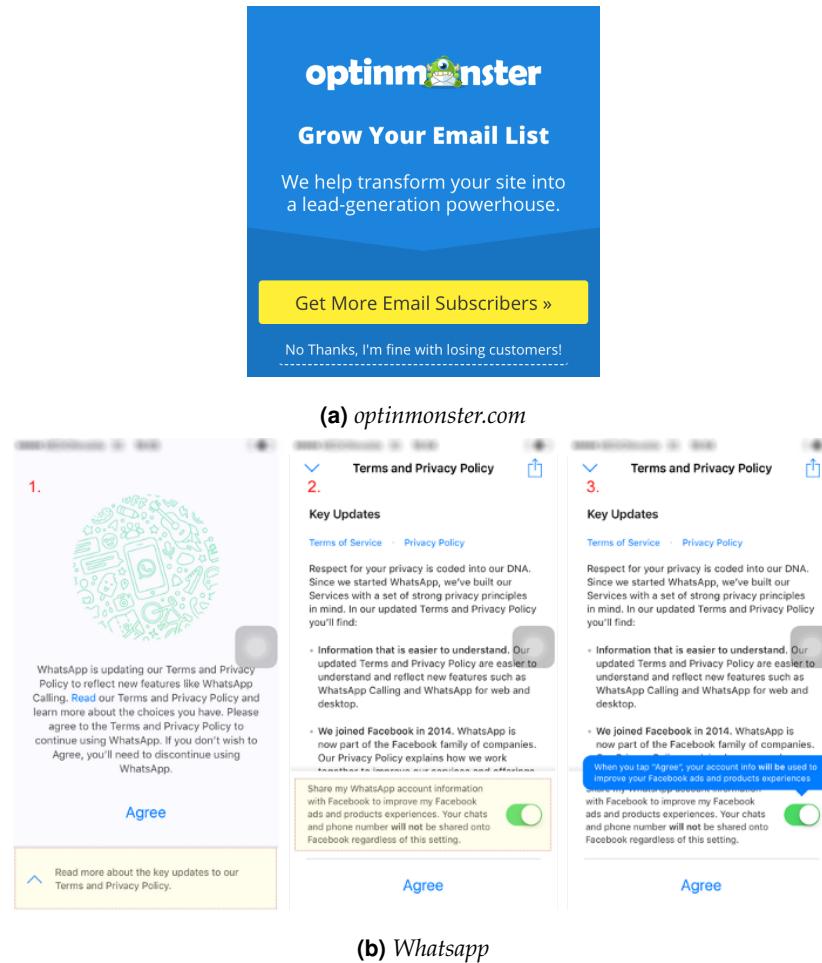


Figura 1.2: Esempi di utilizzo di dark pattern in siti e app

Sul piano legale invece, non esiste ancora una serie di leggi che regola l’implementazione stessa dei dark pattern, anche se, chiaramente, alcuni di essi, per natura, sono considerati illegali; ad esempio con l’introduzione del *Regolamento generale sulla protezione dei dati* (GDPR), molti modelli il cui scopo è ottenere dati dall’utente sono diventati intrinsecamente illegali, perché andrebbero, appunto, a violare il regolamento. Il problema, però, è tutt’altro che risolto, dal momento che tutti quei dark pattern che non violano un regolamento o leggi già esistenti possono essere utilizzati senza problemi (se non etici). A tal proposito la *commissione per il mercato interno e la protezione dei consumatori* (IMCO) del Parlamento europeo si è riunita il 16 Marzo 2022 in un primo incontro con degli esperti (tra cui Harry Brignull stesso) per discutere dell’argomento.

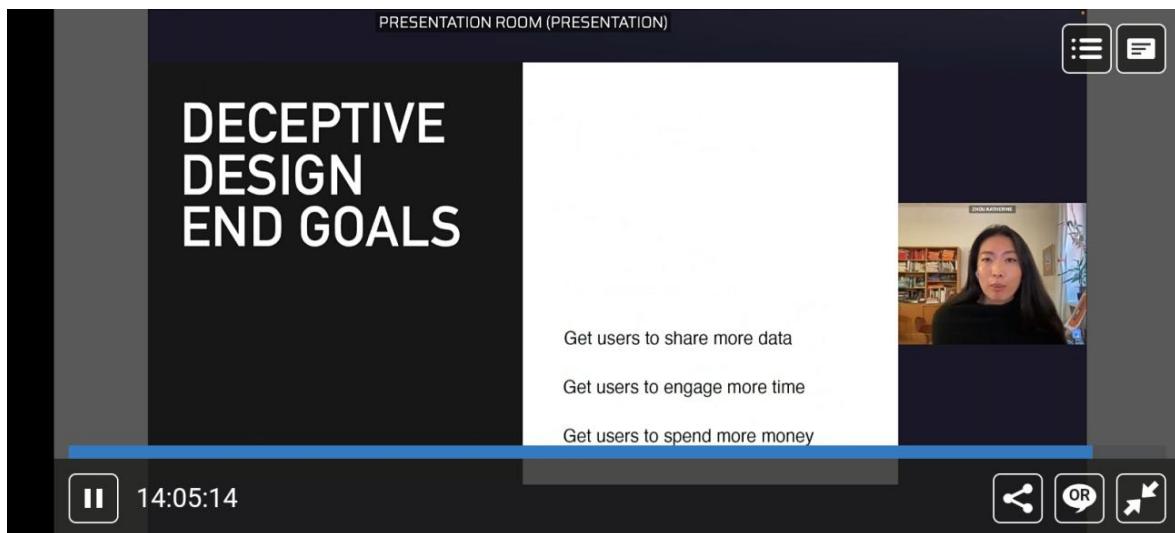


Figura 1.3: Intervento di Katerine Zou il 16 Marzo 2022 alla IMCO

1.2 Motivazioni e Obiettivi

Le motivazioni che hanno incoraggiato questo studio, sono da ricercare nei vari articoli che hanno studiato il fenomeno dei dark pattern; tra i vari, è stato dimostrato infatti da Di Geronimo *et al.* [2020], che molti utenti non sono consapevoli della presenza dei dark pattern nei siti web, e rischiano quindi di caderne vittima.

Gli obiettivi che spingono le varie aziende a inserire dark pattern nei loro prodotti, non si limitano solamente ad un guadagno economico diretto; infatti rispetto ad altri modelli, come ad esempio *sneak into basket* o *hidden costs*, descritti nel Paragrafo 2.1, ci sono dei pattern che mirano ad ottenere dati dagli utenti o aumentare l'engagement degli utenti tramite l'iscrizione a newsletter o inviando notifiche promozionali. Questo tema è stato discusso dalla IMCO, la *Commissione per il mercato interno e la protezione dei consumatori del Parlamento europeo* il 16 Marzo 2022, e in particolare da Katherine Zhou, fondatrice del progetto *Design Ethically*, come si può vedere in Figura 1.3. È quindi necessario che gli utenti siano supportati da un tool che li mette in guardia dai pericoli dei dark pattern, durante la navigazione in internet.

L'obiettivo di questo studio è infatti realizzare un algoritmo che ricerca eventuali istanze di dark pattern che si classificano come *trick question*, in modo che gli utenti possano prestare particolarmente attenzione al testo che gli viene segnalato ed effettuare, di conseguenza, la scelta più conveniente per loro. Le modalità di identificazione, prevedono un approccio *rule-based*, ovvero definendo delle regole che permettono di identificare istanze del pattern tema di questa tesi.

Il suddetto algoritmo potrebbe essere successivamente integrato con Arkan, un sito web che permette, inserendo l'URL di un sito web, di essere informato sulla presenza di tutti i dark pattern, con algoritmi specifici da eseguire per ogni tipologia di modello rilevabile.

1.3 Risultati ottenuti

Analizzando i risultati ottenuti, descritti con più precisione nel Capitolo 4, si può notare come l'algoritmo riesca a ricercare con successo tutte le frasi da analizzare nella pagina, e fornisce buoni risultati per quanto riguarda l'identificazione di *trick question*, però rileva anche molti falsi positivi, principalmente per via di come sono state definite le euristiche; in ogni caso bisogna considerare che future analisi sul tema potrebbero migliorare le euristiche definite in questo studio e quindi, di conseguenza, eventualmente migliorare i risultati.

Inoltre, l'impiego di altri approcci per l'identificazione dei dark pattern, come ad esempio la *computer vision* tramite tecniche di machine learning, potrebbero fornire buoni risultati per quanto riguarda l'identificazione di diversi dark pattern, per la maggior parte quelli che non richiedono interazioni particolari da parte dell'utente, e che sono legati all'estetica delle interfacce, come *sneak into basket*, *hidden costs*, *toying with emotion*, o più in generale quelli che rientrano nella categoria *aesthetic manipulation*, descritti nel Capitolo 2.

Infine, è stato testato l'algoritmo su frasi appartenenti ad altre categorie di dark pattern, per verificare seassegnasse punteggi alti in maniera errata, e non ha dato problemi sotto questo punto di vista, infatti, per quanto riguarda la categoria denotata da Mathur *et al.* [2019] come *activity notification*, simile a *toying with emotions*, l'algoritmo ha dato ottimi risultati: su 250 frasi testate, il punteggio più alto è stato di 0.583; per la categoria *confirmshaming*, invece, il risultato più alto rilevato è stato di 0.33; per le frasi di tipo *misdirection*, invece, associabili a *trick question* e *toying with emotion*, il punteggio più alto è stato di 0.916, assegnato correttamente ad una frase confusa.

1.4 Struttura della tesi

La struttura di questa tesi è descritta di seguito:

1. **Introduzione** - Breve introduzione al mondo dei dark pattern, con alcuni esempi e spiegazione dei problemi che possono causare.
2. **Stato dell'arte** - Descrizione delle varie tassonomie di dark pattern, con vari esempi e un riferimento alle implicazioni sulla privacy. È inoltre presente una classificazione

di dark pattern in base alla possibilità di poterli rilevare automaticamente tramite un algoritmo o meno.

3. **Un approccio rule-based per l'identificazione del dark pattern Trick Question** - Illustrazione della metodologia utilizzata, studio dettagliato del dark pattern *trick question* e descrizione della fase di implementazione dell'algoritmo che lo identifica.
4. **Valutazione preliminare dell'approccio** - In questo capitolo sono riportate le modalità di test dell'algoritmo e i risultati per valutarne l'accuratezza nell'identificazione.
5. **Conclusioni e sviluppi futuri** - Ultimo capitolo contenente le conclusioni e spunti per futuri sviluppi sull'argomento a partire dall'algoritmo, e la sua progettazione.

CAPITOLO 2

Stato dell'arte

2.1 Tassonomia

Il termine "dark pattern" è stato coniato nel 2010 da Harry Brignull, esperto di UX Inglese. La definizione che ha dato a questo termine è "I dark pattern sono modelli di design utilizzati nei siti web e nelle app per indurre gli utenti ad agire contro il proprio reale interesse, come comprare beni o servizi non desiderati o acquistare abbonamenti non voluti". Nello stesso periodo ha anche creato il sito www.darkpatterns.org, all'interno del quale c'è la cosiddetta *Hall Of Shame*, ovvero una lista di siti, segnalati dagli utenti tramite Twitter con l'hashtag *#darkpattern*, che contengono un'istanza di uno o più dark pattern. Invece, nella pagina *Types of Dark Patterns*, è riportata la tassonomia ideata dallo stesso Brignull, che divide i dark pattern in 12 categorie diverse.

Successive ricerche accademiche hanno ampliato questa tassonomia, modificando il significato di determinati modelli o ampliandolo, in base anche all'evoluzione che hanno avuto app e piattaforme web.

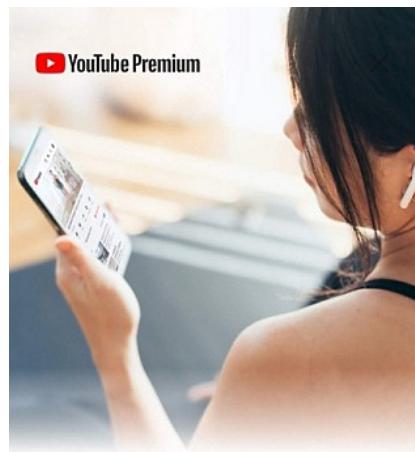
Tra queste, c'è quella di Conti e Sobiesk [2010] in cui si discute di un'altra tassonomia che comprende 11 categorie e 20 sottocategorie; tra queste, compaiono: *distraction* cioè, manipolare lo stile dell'interfaccia, come colori e animazioni particolari, con l'intento di distrarre l'utente e *forced work*, ovvero, costringere l'utente ad effettuare un'operazione per poter utilizzare una specifica funzionalità, che altrimenti è inaccessibile o disattivata, molto spesso utilizzato per la visualizzazione di annunci pubblicitari. Nel 2016, Bösch *et al.* [2016] hanno presentato una

classificazione di otto “Dark Strategies”, discutendo di come alcune piattaforme online e siti web utilizzano i dark pattern per danneggiare la privacy dell’utente, usando, per esempio, un linguaggio fuorviante che fa credere all’utente di trarre beneficio da determinate scelte sulla condivisione dei propri dati, o non rivelando dettagli che hanno un impatto rilevante sulla privacy, in risposta al lavoro di Hoepman [2014]. Tra le “Dark Strategies”, appunto, figurano *forced registration* cioè forzare la registrazione di un account per accedere a determinate funzionalità, inaccessibili altrimenti, e *hidden legalese stipulations* ovvero creare dei *termini e condizioni* volutamente lunghi e con formulazioni confusionarie per nascondere informazioni; il linguaggio estremamente formale e i testi molto lunghi utilizzati in molte politiche sulla privacy diventa una sorta di "dark pattern" dal momento che potrebbe impedire all’utente di capire quello che sta leggendo, oppure sfavorirne la lettura del tutto.

La tassonomia più recente è stata proposta da Gray *et al.* [2018] raccogliendo un insieme di 118 siti web che contengono istanze di alcuni dark pattern, riportati su social media, blog di esperti nel settore UX e dalle loro esperienze con prodotti della vita di tutti i giorni. Lavorando poi sulla tassonomia originariamente presentata da Brignull, hanno aggiunto categorie e sottocategorie e adattato ed esteso il significato di alcuni DP tenendo conto dell’evoluzione che dal 2010 fino ad oggi hanno avuto le piattaforme online e soprattutto del target di utenti a cui esse sono rivolte. In particolare Gray *et al.* [2018] hanno proposto 5 categorie di dark pattern: *nagging, obstruction, sneaking, interface interferences, aesthetic manipulation*.

2.1.1 Nagging

Questa categoria include tutte quelle interruzioni durante l’esecuzione di un task da parte dell’utente. In altre parole si tratta dell’interruzione improvvisa (una o più volte) del task che l’utente sta svolgendo, tramite un altro task che non è correlato a quello originale. *Nagging* può presentarsi come pop-up, dialog, o pubblicità improvvise a schermo intero che oscurano l’interfaccia e che quindi spostano il focus dell’utente mentre sta effettuando un’operazione. Nell’esempio in Figura 2.1, si può notare come l’app di Youtube per dispositivi mobile, mostra una pubblicità a schermo intero, verosimilmente all’avvio, quando l’utente si aspetterebbe di vedere la schermata di home; inoltre non c’è nessun controllo che permette all’utente di abbandonare la schermata.



**Don't miss the perks of
Premium**

Try YouTube Premium. We'll email you seven days before your trial ends.

1 MONTH FREE

Figura 2.1: Esempio del dark pattern *nagging* nell'app *Youtube*

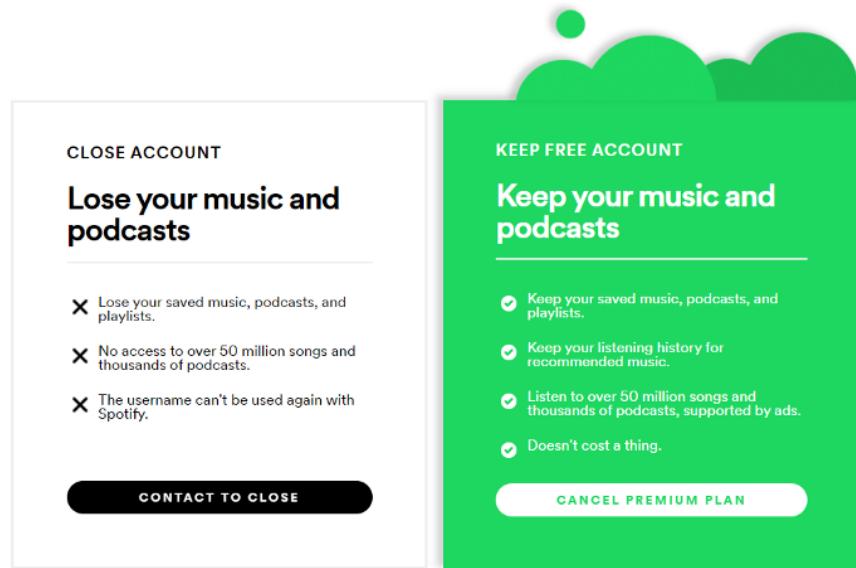


Figura 2.2: Esempio del dark pattern *roach motel* sul sito *Spotify*

2.1.2 Obstruction

Questa categoria raggruppa tre tipi di dark pattern: *roach motel*, *price comparison prevention* e *intermediate currency* tutti e tre, originariamente definiti da Brignull; rientrano in questa categoria, tutti quei modelli che rappresentano un ostacolo inserito nel mezzo di un task flow, con l'intento di scoraggiare l'utente a completare quella particolare operazione. In altre parole Gray *et al.* [2018] definiscono questa categoria come "rendere un processo più difficile di quanto debba essere, per dissuadere l'utente dal compiere una determinata azione".

Roach Motel

Originariamente Brignull raggruppa in questa categoria "tutte quelle situazioni in cui risulta facile per l'utente entrarne, ma difficile uscirne (per esempio iscriversi/disiscriversi da un abbonamento)" [Brignull, 2010]. Spesso questo modello si rileva quando, su una piattaforma web, un utente riesce facilmente a registrarsi, ma risulta poi difficile cancellare l'account, o addirittura impossibile.

Nell'esempio in Figura 2.2, sul sito di Spotify si nota immediatamente, che per cancellare definitivamente l'account (l'opzione a sinistra), bisogna cliccare sul bottone *contact to close*, che rimanda ad un form per parlare con l'assistenza clienti, per cui l'utente non può cancellare l'account in completa autonomia.

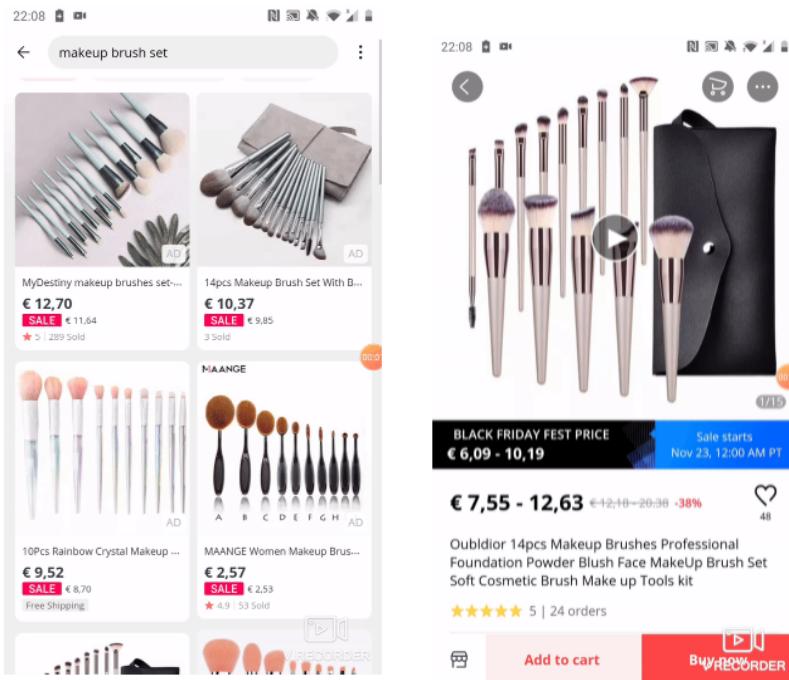


Figura 2.3: Esempio del dark pattern *price comparison prevention* sul sito AliExpress

Price Comparison Prevention

Questo modello è utilizzato, molto spesso, per evitare che l’utente confronti i prezzi di beni o servizi simili, sullo stesso sito, o tra siti diversi. A questo scopo i siti web utilizzano svariate tecniche come, per esempio, fornire poche informazioni sul prodotto, oppure impedire all’utente di copiare il nome del prodotto che sta visualizzando, così da rendere più difficile cercarlo altrove; altre volte, invece, vengono offerti piani di iscrizione a pagamento con diverse caratteristiche, ma non viene concessa la possibilità di confrontare facilmente i prezzi dei diversi piani, per metterli in relazione con i servizi offerti.

Nell’esempio in Figura 2.3, nell’app AliExpress, l’utente può effettuare una query per cercare prodotti; i risultati della query effettuata, vengono mostrati in una lista nella quale ad ogni prodotto è associato un prezzo singolo, ma, cliccando su un prodotto, si scopre che in realtà quel prezzo cambia a seconda delle varianti (colore, taglia...) e quindi la lista iniziale risulta essere fuorviante perché non permette di prendere una decisione corretta nel confrontare i prezzi.

Intermediate Currency

È un altro dark pattern, sottotipo di *obstruction*, molto popolare soprattutto nei videogiochi, che prevede l’utilizzo di una valuta virtuale per effettuare acquisti nel gioco; per esempio

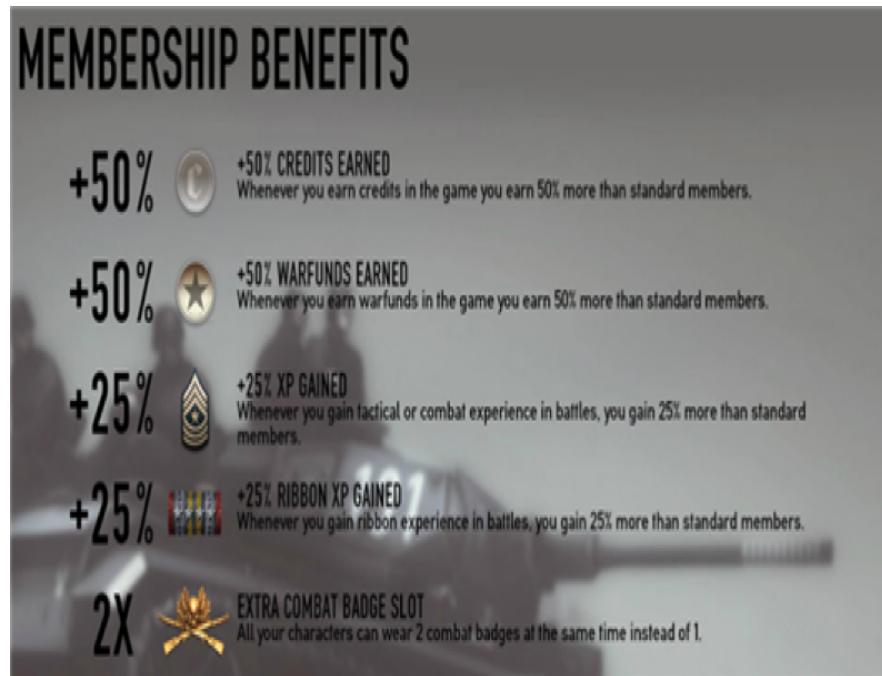


Figura 2.4: Esempio del dark pattern *intermediate currency* nel gioco *Heroes and Generals*

l’utente può comprare potenziamenti tramite dei "diamanti" (valuta virtuale), e può ottenere questi "diamanti" acquistandoli con soldi reali. Questo pattern mira a confondere l’utente sul valore che i soldi reali hanno all’interno del gioco; congiuntamente alla valuta virtuale spesso le app usano anche formulazioni confusionarie sui benefici che si ottengono dai potenziamenti (usando per esempio valori in percentuale anziché valori fissi).

Nell’esempio della Figura 2.4, il DP è implementato nel gioco *Heroes and Generals*: tramite soldi reali è possibile comprare "gemme", tramite le quali, a loro volta, è possibile ottenere dei benefici e potenziamenti; si nota facilmente l’utilizzo delle percentuali sulla sinistra dell’elenco; in questo modo, non usando dei valori fissi, è più difficile per l’utente relazionare il costo (in soldi reali) di un beneficio all’incremento delle statistiche che può ottenere.

2.1.3 Sneaking

Questa categoria comprende i DP *forced continuity*, *hidden costs*, *sneak into basket* e *bait and switch*, tutti DP originariamente definiti da Brignull nel 2010. Sono racchiuse in questa categoria tutti pattern che cercano di nascondere o ritardare la divulgazione di informazioni importanti per l’utente, in modo da influenzare le decisioni che prenderà; infatti le informazioni che vengono nascoste o comunicate in ritardo permetterebbero all’utente di non effettuare azioni indesiderate, se fosse stato a conoscenza di esse fin da subito.

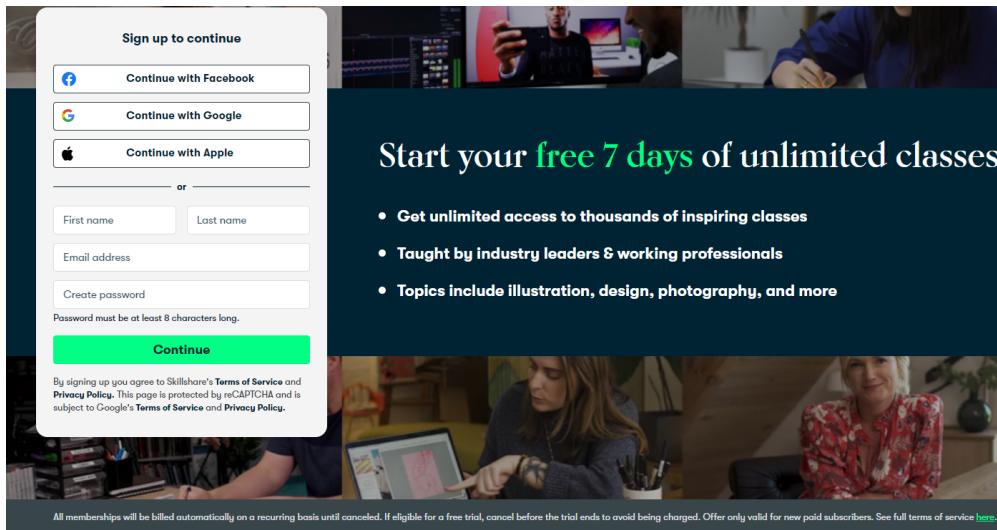


Figura 2.5: Esempio del dark pattern *forced continuity* nel sito *skillshare.com*

Forced Continuity

Questo DP permette ad un sito web di continuare ad addebitare all’utente il costo di un abbonamento per un servizio, anche dopo la sua scadenza. Per esempio si può verificare questa situazione quando l’utente si iscrive per utilizzare il periodo di prova gratuita di una piattaforma, ma al termine di questo periodo, dalla sua carta di credito, viene automaticamente detratto il costo dell’abbonamento per un certo periodo di tempo, anche se non richiesto. Nell’esempio in Figura 2.5 il servizio propone un periodo di prova gratuito di 7 giorni, ma leggendo il messaggio piccolo in basso, si scopre che dopo 7 giorni il costo del servizio viene addebitato automaticamente se non viene interrotto manualmente.

Hidden Costs

Questo DP consiste nel pubblicizzare un prodotto ad un determinato prezzo, ma al momento dell’acquisto vengono aggiunti costi aggiuntivi, come tasse o commissioni, o alternativamente, costi di spedizione elevati.

Nell’esempio in Figura 2.6, il sito *broadway.com* permette di acquistare dei biglietti per uno spettacolo, al prezzo di \$ 59.90 ognuno; al momento del pagamento, però, vengono aggiunti ulteriori \$ 14.88 per ogni biglietto, per un servizio che il sito chiama *service & handling*, ovvero delle tasse di servizio o commissioni, confondendo l’utente che fino a un attimo prima pensava di spendere meno.

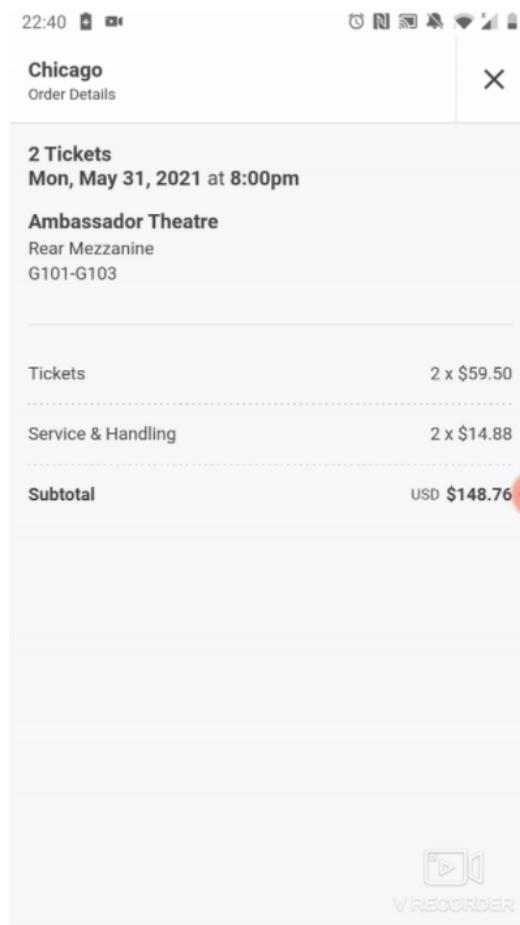


Figura 2.6: Esempio del dark pattern *hidden costs* nel sito *broadway.com*

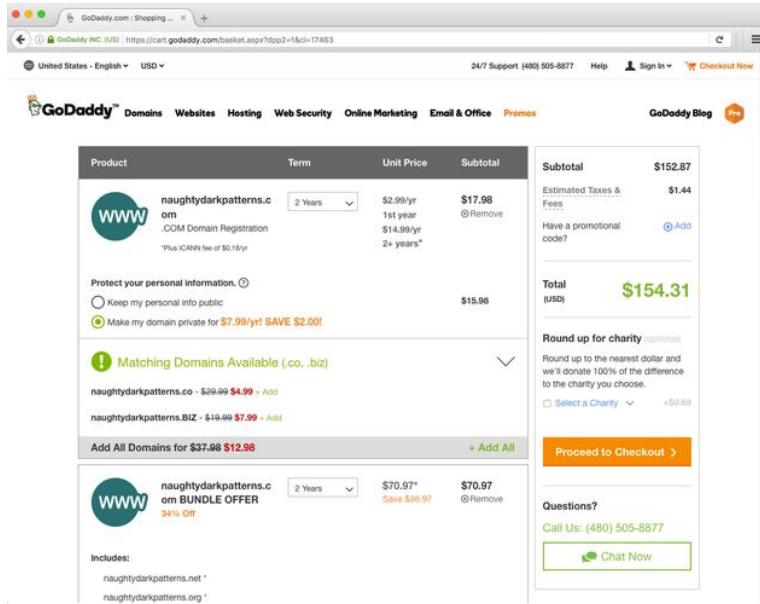


Figura 2.7: Esempio del dark pattern *sneak into basket* nel sito godaddy.com

Sneak into Basket

Questo modello è di popolare utilizzo soprattutto nei siti web di e-commerce: nel momento in cui l'utente aggiunge un prodotto al suo carrello virtuale, vengono inseriti al suo interno, anche dei prodotti che non ha richiesto, la cui presenza è giustificata spesso da suggerimenti o articoli correlati; questo pattern può indurre l'utente ad acquistare involontariamente gli articoli aggiunti dal sistema se non rimossi manualmente, assumendo che egli faccia attenzione e li noti.

Nell'esempio in Figura 2.7 il pattern è rilevato nel sito *godaddy.com*: in particolare acquistando un dominio, nella figura al prezzo di \$ 17.98, il sito aggiunge automaticamente al carrello un pacchetto che contiene un insieme di domini correlati a quello che si vuole (intenzionalmente) acquistare, facendo arrivare il totale da pagare a \$ 154.31

Bait and Switch

Questo pattern modifica il comportamento di un'azione effettuata dall'utente, rendendola svantaggiosa per quest'ultimo. Si verifica ad esempio quando si cerca di chiudere una pubblicità con il bottone "X", ma invece si viene rimandati ad un sito esterno.

Nell'esempio in Figura 2.8, il sito *reddit.com* mostra una serie di post in una lista verticale, ognuno dei quali ha nome dell'autore, data di pubblicazione e titolo del post; cliccando su uno dei post, esso si "apre" mostrando anche una breve descrizione fornita dall'autore. In ogni caso, però, c'è un elemento nella lista che è etichettato come *promoted*, quindi una pubblicità,

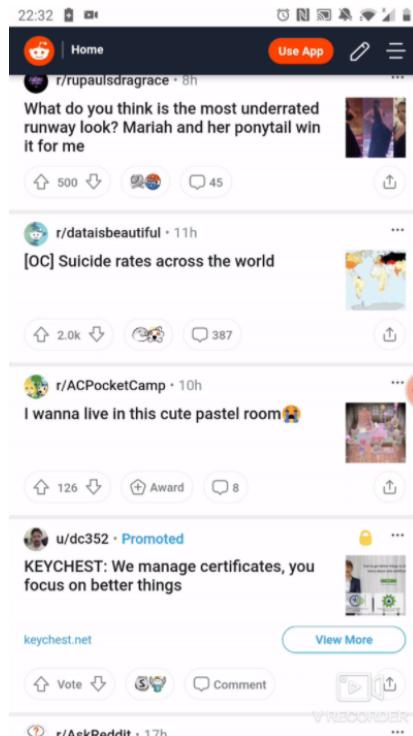


Figura 2.8: Esempio del dark pattern *bait and switch* nel sito *reddit.com*

il cui design è però identico agli altri post, e cliccando su di esso si viene reindirizzati al sito dello sponsor.

2.1.4 Interface Interferences

Questa categoria comprende una serie di design volutamente fuorvianti dell’interfaccia che non permettono all’utente di essere a conoscenza di informazioni importanti, in modo da fargli compiere delle operazioni rispetto ad altre, nascondendo o mascherandone alcune. Questa categoria include 4 DP: *hidden information*, *preselection*, *aesthetic manipulation* e *trick questions*.

Hidden Information

Questo DP è utilizzato per nascondere informazioni all’utente attraverso dettagli di design come rendere un testo grigio come se fosse disabilitato, o dello stesso colore del background di una pagina per nasconderlo completamente, oppure di piccole dimensioni. Con questa tecnica si cerca di far passare un’informazione rilevante come irrilevante per l’utente.

Nell’esempio in Figura 2.9 nel sito *rac.co.uk*, viene nascosta – e preselezionata – una checkbox per iscriversi alla newsletter in un menu nascosto, che l’utente deve aprire manualmente, cliccando *more info* per visualizzarlo, per poi deselectare l’opzione, se la nota.

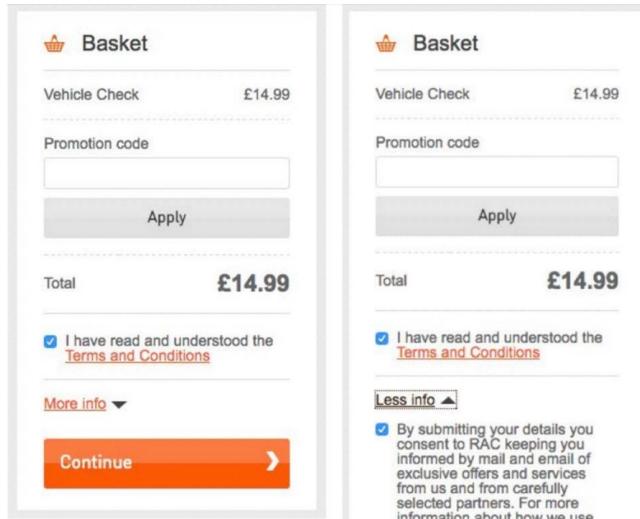


Figura 2.9: Esempio del dark pattern *hidden information* nel sito *rac.co.uk*

Preselection

Questo DP può essere rilevato in tutte quelle situazioni in cui l’utente viene messo di fronte a varie scelte, ma una risulta, come suggerisce il nome, preselezionata automaticamente dal sistema; spesso questo modello è realizzato con delle checkbox, e quella preselezionata è la scelta che il proprietario del sito vuole che l’utente faccia, anche se contro il suo interesse, come per esempio iscriversi a una newsletter, o dare il consenso per la condivisione dei suoi dati personali.

In Figura 2.9 si può notare la checkbox per ricevere annunci promozionali tramite email preselezionata dal sito e nascosta sotto un menù a tendina; il DP consiste nel fatto che per evitare di ricevere email, bisogna manualmente deselezionare l’opzione, comportamento che l’utente non si aspetta di default.

Aesthetic Manipulation

Questa sottocategoria comprende 3 DP: *disguised ad*, *false hierarchy* e *toying with emotion*. Tutti questi modelli riguardano più l’aspetto che le funzionalità di app e siti web; sono cioè scelte di design accuratamente prese, per spostare l’attenzione dell’utente, o deviare le sue convinzioni.

Toying with emotions

Questo DP è implementato per influenzare l’utente nel compiere una specifica operazione, facendogli provare determinate emozioni, tramite elementi come formulazioni di frasi

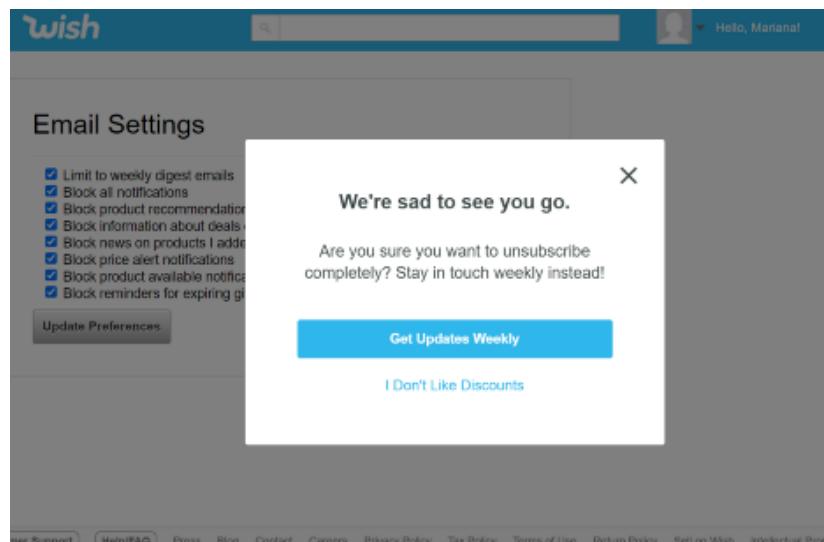


Figura 2.10: Esempio del dark pattern *toying with emotion* nel sito *wish.com*

particolari, colori, o immagini. In Figura 2.10, questo pattern è in azione sul sito *wish.com*: se l’utente vuole disiscriversi dalla newsletter, il sito, con la giustificazione di inviare sconti e promozioni tramite email, chiede all’utente se vuole confermare la sua scelta, ma l’opzione per confermare è etichettata come *non mi piacciono gli sconti*, in modo da creare un senso di “colpevolezza” nell’utente.

False Hierarchy

Questo DP consiste nell’influenzare l’utente facendogli effettuare un’azione piuttosto che un’altra, facendo in modo che l’azione “voluta” sia più ovvia di un’altra, a livello grafico, tramite colori più accesi o animazioni.

Nell’esempio in Figura 2.11, il modello è rilevato sul sito *lastminute.com*, e in particolare si nota come il bottone per accettare e continuare ha uno sfondo colorato, che lo rende molto evidente, mentre a sinistra, l’opzione per rifiutare è un normale link senza sfondo, decisamente meno appariscente.

Disguised Ad

Questo DP è utilizzato largamente da siti web e app, per camuffare le pubblicità, in modo da confondere l’utente, o in modo da attirare l’attenzione su di esse; nei casi più estremi, l’intera pagina è coperta da un link invisibile, e quindi ovunque si effettua un click nella pagina, si viene rimandati al sito esterno dello sponsor.

In Figura 2.8, si può notare come il sito *reddit.com*, nella sua versione mobile, mostri nella

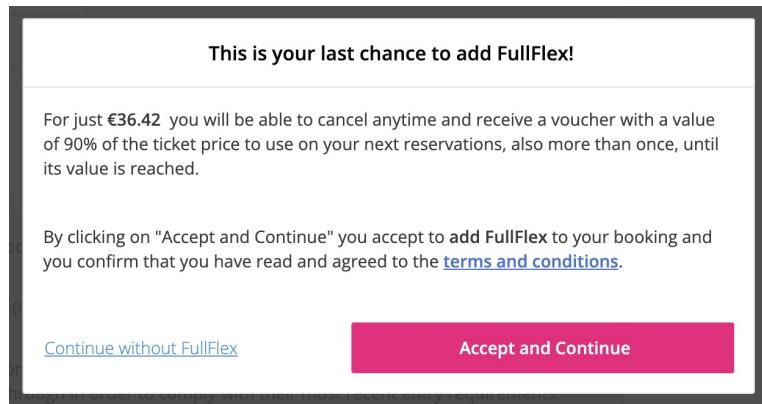


Figura 2.11: Esempio del dark pattern *false hierarchy* nel sito *lastminute.com*

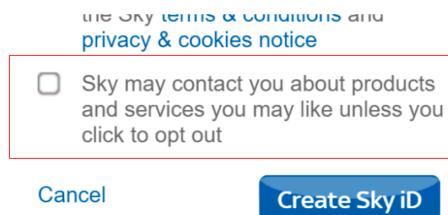


Figura 2.12: Esempio del dark pattern *trick questions* nel sito *sky.com*

home una lista di post, però uno di questi in realtà è una pubblicità, ed è etichettato come *promoted*, infatti cliccando si viene rimandati al sito dello sponsor che ha comprato l'annuncio.

Trick Questions

Questo DP è utilizzato con l'intento di confondere l'utente quando legge un'informazione, con frasi formulate in maniera confusa, o con doppie negazioni, per manipolare la volontà dell'utente.

Nella Figura 2.12 si può notare un esempio in cui viene implementato questo pattern: leggendo il testo si capisce che l'utente deve **cliccare** (azione positiva) per **rifiutare** (azione negativa) di essere contattato tramite email per promozioni e offerte; la frase cerca dunque di generare un carico cognitivo nell'utente, per confonderlo.

2.1.5 Forced Action

Questa categoria include una serie di dark pattern implementati per impedire all'utente di accedere ad una funzionalità, se prima non ha eseguito un altro task, spesso svantaggioso per lui, mascherato come un'azione da cui ne trae un forte beneficio.



Figura 2.13: Esempio del dark pattern *social pyramid* nell'app *Farmville*

Social Pyramid

Questo DP è prevalentemente usato nei social network, e forza un utente ad invitare altre persone ad utilizzare il sito o l'app, in cambio dell'accesso a funzionalità che sarebbero non utilizzabili altrimenti. Anche nei videogiochi è molto diffuso, e incentiva l'utente tramite dei potenziamenti o bonus che gli vengono regalati invitando amici.

In Figura 2.13, questo pattern è rilevato nell'app *Farmville*: il gioco impedisce di accedere a determinate funzionalità, se non si invitano prima amici ad iscriversi.

Privacy Zuckering

Questo DP, inganna l'utente, in modo da fargli condividere più dati personali di quanti ne voglia fornire in realtà. Lo scopo di questo modello è, solitamente, trarre profitto dai dati dell'utente vendendoli ad aziende di terze parti; per "proteggersi" legalmente, di solito la clausola che autorizza al trasferimento dei dati personali viene inclusa nei termini di utilizzo, che devono essere obbligatoriamente accettati.

Nell'esempio in Figura 2.14, un'istanza del modello è rilevata nell'app Messenger di Facebook: quando si avvia dopo la prima installazione, l'app chiede il permesso di accedere alla rubrica dell'utente per creare una rete di amici. In questo modo l'app può utilizzare la rete di amici per fornire pubblicità personalizzate. Questa funzione di per sé potrebbe essere anche utile, e non oscura, tuttavia il modo in cui viene chiesto il consenso lo è: per negare l'accesso, infatti,

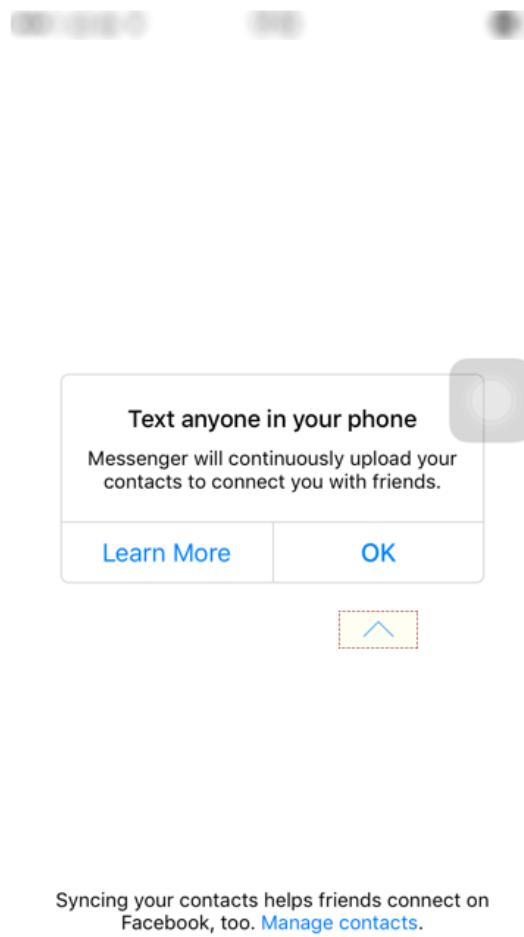


Figura 2.14: Esempio del dark pattern *privacy zuckering* nell'app Messenger

non c'è un tasto esplicitamente etichettato con "no", ma, bisogna cliccare su *learn more*, leggere la privacy policy integralmente, e infine negare il consenso. È da notare anche, la freccia che punta verso la scelta *OK*, come a indicare che quella è la scelta giusta (caso di *false hierarchy*).

Gamification

Questo DP incentiva l'utente ad effettuare ripetutamente determinati task, promettendo un "premio" in cambio. Questo pattern, a livello cognitivo è molto influente, perché sfrutta il *sistema di ricompensa* del cervello: Hamari *et al.* [2014] hanno dimostrato che il cervello umano è particolarmente suscettibile a questo sistema, dal momento che se si fallisce nel completare un task, e quindi non si ottiene il premio, il cervello risponde negativamente (generando *microstress*).

Nell'esempio in Figura 2.15, si nota, l'app *Candy Crush Saga* rende dei livelli impossibili da completare se non si invitano amici o senza comprare vite aggiuntive a pagamento.

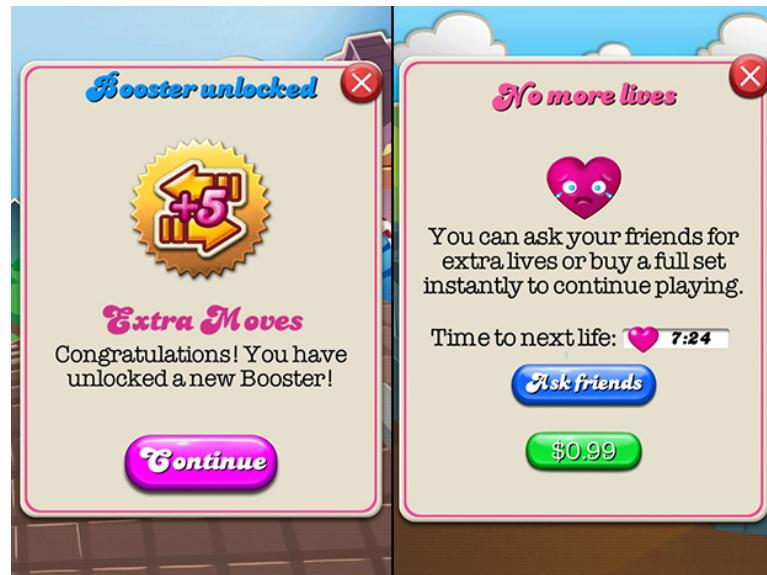


Figura 2.15: Esempio del dark pattern *gamification* nell'app *Candy Crush Saga*

2.2 Impatto dei Dark Pattern

I DP sono particolarmente efficaci nel loro (oscuro) intento, come dimostrato da Di Geronimo *et al.* [2020] tramite uno studio condotto su un campione di utenti. Questo accade perché ormai i team di sviluppo delle interfacce utente sono composti non solo da informatici, ma anche da psicologi, sociologi e antropologi; sfruttando le conoscenze di questi ultimi dunque, il lavoro di sviluppo del design è condotto tenendo conto dei bias cognitivi. Proprio per questo motivo molto spesso chi si trova davanti a un dark pattern, non ha piena coscienza della presenza di quest'ultimo. In particolare, lo studio di Di Geronimo *et al.* [2020] si concentra sulla percezione che l'utente ha dei DP: tramite un sondaggio online hanno determinato una statistica riguardante la percentuale di utenti che sono in grado di riconoscere un DP; è stato chiesto, a 589 partecipanti, di guardare un video di 30 secondi in cui viene utilizzata un'app e successivamente di specificare se avessero rilevato un design "cattivo". Il 55% degli utenti ha risposto di non aver individuato alcuna anomalia nel design, confermando dunque che la maggior parte di chi utilizza un sito web o un'app non è in grado di riconoscere la presenza di un DP; al restante 45% è stato chiesto inoltre di dichiarare che anomalie avevano riscontrato, e solo il 7% di essi ha esplicitamente detto di aver individuato un dark pattern. Inoltre è esplicitamente dichiarato da Di Geronimo *et al.* [2020] che dato che la maggior parte degli utenti non è in grado di rilevare la presenza dei dark pattern, sono necessari degli strumenti automatici che supportino l'utente nel riconoscimento della presenza di design oscuri, per esempio attraverso un livello di rischio indicante la possibilità che un determinato

sito includa uno o più dark pattern.

2.3 Implicazioni sulla privacy

I DP assumono una rilevanza elevata anche in relazione alle tematiche di privacy: alcuni modelli infatti vengono utilizzati per ottenere il consenso dall'utente di trattare più dati personali di quanti esso ne intenda fornire, infatti i DP *privacy zuckering*, *trick question* e *hidden information* possono essere tutti impiegati a questo scopo. Le informative per la privacy, per esempio, sono presenti in tantissimi siti web, e possono spesso essere considerate una sorta di dark pattern per via del linguaggio formale utilizzato, che ne scoraggia, in un certo senso, la lettura; i fornitori dei servizi hanno un forte interesse nel poter "lavorare" con i dati personali, sia per proporre pubblicità personalizzate, che per cederli ad aziende di terze parti. Per questo motivo i DP, ultimamente, sono stati oggetto di attenzione da parte di governi ed istituzioni, come la commissione europea, nell'ambito della tutela dei consumatori, o lo stato della California, negli USA. Nel 2018 infatti è stato proposto il *California Consumer Privacy Act* (CCPA) da un gruppo noto come *Californians for Consumer Privacy*, sottoforma di petizione, diventata poi una normativa vera e propria il 28 Giugno 2018, e diventata effettiva il 1 Gennaio 2020. Nello specifico questo documento disciplina il modo in cui le aziende **di tutto il mondo**, sono autorizzate a trattare i dati dei cittadini residenti in California, attenuando quindi gli effetti dei DP quando vengono usati per ottenere il consenso al trattamento delle informazioni personali. Dev'essere garantito infatti, ai cittadini californiani, il diritto di opporsi alla vendita dei propri dati ad aziende terze, nonché la comunicazione dei dati raccolti e la loro cancellazione.

Anche l'Unione Europea si sta muovendo in questo verso tramite l'approvazione del *Digital Service Act*, un documento che stabilisce nuove norme per i fornitori di servizi online, relativamente al trattamento dei dati.

È da notare, in ogni caso, che sia il CCPA, che il DSA, non vietano direttamente l'utilizzo di specifici DP, in ogni caso, però, la commissione europea sta organizzando degli incontri per discutere di queste tematiche.

D'altra parte, con un tool che indica all'utente un livello di rischio per la presenza dei DP, l'interazione con i siti web potrebbe avvenire diversamente, facendo più attenzione alle scelte che vengono effettuate rispetto alla condivisione dei propri dati; ad esempio prima di registrarsi ad un sito, l'utente potrebbe far "analizzare" il sito ad un programma, che fornisce un indice di rischio rispetto alla presenza di potenziali pericoli per la privacy.

2.4 Identificazione dark pattern

Non tutti i DP possono essere identificati tramite un algoritmo, per via della loro natura eterogenea. Tuttavia alcuni di essi sono implementati in un modo che consente di identificarli *automaticamente* attraverso l'esecuzione di una serie di passaggi. Nella Tabella 2.1 sono listati i dark pattern, insieme alla possibilità di poterli rilevare automaticamente attraverso un algoritmo, e le motivazioni riguardo, appunto, questa possibilità. In particolare questa classificazione è stata redatta insieme a questo studio, nel contesto del corso dell'esame *Interazione Uomo-Macchina* e delle attività di tirocinio; il processo di raggruppamento è stato inoltre svolto in maniera collaborativa insieme ad altri tre studenti che, individualmente analizzavano definizioni ed esempi dei vari dark pattern per poi confrontarsi e procedere alla catalogazione.

DP	Identificazione	Motivazione
Obstruction		
Roach Motel	Automatica	Può essere identificato cercando all'interno delle pagine web bottoni con diciture come "Elimina account" oppure "Interrompi abbonamento"
Price Comparison Prevention	Automatica	Analizzando gli script presenti nella pagina web, si potrebbe stabilire se è stata disattivata la funzionalità di "copia e incolla"; questo metodo, seppur non permette l'identificazione di tutte le istanze di questo dp, è comunque sufficiente per rilevarne un vasto sottoinsieme, dal momento che è l'implementazione più comune.
Intermediate Currency	Automatica	Ricercando una serie di keywords all'interno della pagina come "gemme", "monete" o "diamanti", soprattutto nel contesto di un acquisto è facile rilevare la presenza di questo pattern.
Sneaking		
Hidden Costs	Automatica	Simulando l'acquisto di un bene o un servizio, questo dark pattern potrebbe essere identificato in una pagina web, confrontando il prezzo all'inizio della sessione, con quello dichiarato al momento del pagamento.
Sneak into Basket	Automatica	Come Hidden Costs anche in questo caso è sufficiente simulare l'aggiunta di un prodotto al carrello per poi contare il numero di prodotti effettivamente presenti in esso.
Forced Continuity	Manuale	Questo dark pattern può essere individuato solo manualmente, determinando se dopo il periodo di prova gratuita di un servizio sono stati addebitati costi automaticamente.
Bait and Switch	Manuale	Non è possibile identificare automaticamente questo pattern dal momento che non è possibile sapere a priori quale sarebbe la normale interazione di un comando che viene modificata per ingannare l'utente.
Interface Interference		
Toying with emotions	Automatico	Dal momento che questo dp è implementato esclusivamente tramite testo, con algoritmi di NLP (Natural Language Processing) può essere identificato.
Disguised Ad	Automatico	Questo pattern potrebbe essere identificato ricercando in una pagina dei bottoni o altri controlli che reindirizzano a siti esterni, effettuando poi le opportune valutazioni.
Trick Question	Automatico	Anche questo è un modello che può essere identificato tramite applicazione di algoritmi di NLP.
Preselection	Automatico	Per identificare questo dark pattern si può procedere analizzando le pagine web cercando le checkbox presenti già selezionate al caricamento della pagina.
Hidden Information	Manuale	Dato che questo dark può essere implementato in maniere molto diverse può essere rilevato solo da un giudizio umano.
False Hierarchy	Manuale	In maniera simile al dark pattern precedente anche questo può essere implementato in molteplici modi che non ne consentono l'identificazione tramite un algoritmo.
Forced Action		
Social Pyramid	Manuale	Anche questo pattern può essere implementato in modalità diverse e quindi non può essere rilevato in maniera automatica.
Privacy Zuckering	Manuale	Analogamente a Social Pyramid anche questo pattern non può essere rilevato tramite un algoritmo per via degli svariati modi in cui può essere implementato.
Gamification	Manuale	Date le modalità di implementazione anche questo dark pattern non è identificabile, se non manualmente.

Tabella 2.1: Tabella di riepilogo riguardo l'identificazione dei dark pattern

CAPITOLO 3

Un approccio rule-based per l'identificazione del dark pattern Trick Question

Come detto in precedenza, il dark pattern *trick question* viene implementato tramite delle domande da porre all'utente. Il "trucco" consiste nel formulare la domanda in modo confuso; così facendo, leggendo la frase velocemente, essa sembra avere un determinato significato, ma in realtà, leggendola più attentamente, ne ha uno totalmente diverso (o opposto). Per ottenere questo risultato molto spesso vengono utilizzate formulazioni confusionarie, doppie negazioni o un linguaggio che in altri modi manipola le intenzioni dell'utente. Alcune volte vengono sfruttati dei bias cognitivi, ad esempio associare un azione positiva ad un effetto negativo, o viceversa; un esempio di quest'ultima tecnica si può trovare nella figura 2.12: leggendo attentamente si nota che per rifiutare l'iscrizione alla newsletter, l'utente deve selezionare la checkbox; è ovvio che questa frase vuole confondere l'utente, che è abituato a deselezionare una casella di controllo per declinare il consenso. Principalmente questo tipo di dark pattern viene utilizzato per disorientare l'utente e fargli compiere delle scelte che normalmente non avrebbe compiuto.

L'identificazione di questo dark pattern si configura, quindi, come un problema di analisi del linguaggio naturale (*Natural Language Processing*). Nel contesto dell'elaborazione del linguaggio naturale, esistono molti algoritmi già collaudati che vengono messi a disposizione da alcune librerie; alcuni di questi sono *P.O.S. Tagging*, *stemming* e *sentiment analysis*. Tuttavia è l'impiego simultaneo di alcuni di queste funzioni che ha permesso di implementare un meccanismo per l'identificazione di questo dark pattern; il motivo della scelta di impiegare

diversi algoritmi di *NLP* contestualmente, è scaturito dal fatto che le frasi che si configurano come *trick question* non seguono regole precise e quindi, ad esempio, non è sufficiente ricercare doppie negazioni in un testo per concludere che esso è stato reso volontariamente confuso. Per risolvere questo problema, è stato analizzato un dataset di dark pattern, alla ricerca di pattern comuni tra le frasi che rientrano nella categoria *trick question*, in modo da stabilire delle euristiche. Nello specifico il dataset di riferimento è quello rilasciato da Mathur *et al.* [2019], che, però, contiene solo nove frasi identificate come *trick question*; in ogni caso, quindi, l'identificazione di questo specifico dark pattern è ancora più "challenging".

In questo capitolo viene illustrato il processo di implementazione dell'algoritmo a partire dalla ricerca del dataset, fino alla stesura del codice.



Figura 3.1: Illustrazione metodologia utilizzata

3.1 Analisi del dataset

Come si può notare in Figura 3.1, la prima fase è stata ricercare un dataset di DP, in modo da avere delle frasi da analizzare successivamente; questa fase si è conclusa, come detto in precedenza tramite delle ricerche sul web, che hanno portato al dataset creato da Mathur *et al.* [2019], del quale si può trovare un estratto in Tabella 3.1.

L'approccio utilizzato per analizzare il dataset è un metodo di ricerca detto *Grounded Theory*, secondo il quale osservazione ed elaborazione teorica devono avvenire contemporaneamente; questa tecnica quindi pone il focus sui dati e non sulle teorie, che derivano direttamente dall'analisi dei dati. Tra le altre cose, nell'utilizzare la *GT* proprio per la sua natura, bisognerebbe ignorare qualsiasi prestrutturazione teorica.

Durante l'analisi delle frasi contenute nel dataset si è cercato di identificare dei pattern comuni tra loro; uno di questi è sicuramente la lunghezza: tutte, infatti, sono formate da più di 10 parole, ed effettivamente frasi lunghe sono inevitabilmente più confuse. Un altro fattore comune a tutti i periodi, è la presenza, talvolta anche ripetuta, delle parole **do not** o, in forma

contratta, **don't**; la presenza di queste negazioni, in delle frasi che nella maggior parte dei casi saranno posizionate adiacenti a delle checkbox, fa pensare alla volontà, da parte dei designer, di voler confondere l'utente associando un'azione positiva (cliccare) per **negare** un consenso. Durante la terza fase, sono state definite le euristiche da impiegare per l'algoritmo, assegnando ad ognuna di esse un "peso", a valle di sperimentazioni in cui sono state valutate le prestazioni dell'algoritmo rispetto ai valori assegnati.

Le regole definite sono descritte di seguito:

E1. La frase contiene "if you", "don't" o "do not".

Come detto in precedenza tutte le frasi presenti nel dataset contengono almeno una di queste parole; il peso assegnato a questa regola è di 2.

E2. La frase contiene almeno tre delle seguenti parole: "box", "tick", "uncheck", "receive", "email", "newsletter", "checkbox", "please", "discount", "benefit", "offer", "promotion", "receive", "special offer", "unsubscribe".

Queste parole sono molto ricorrenti nelle frasi del dataset, e riguardano temi per i quali di solito le aziende hanno un interesse particolare, ad esempio *newsletter*, perché aumentando il numero di utenti iscritti ad una mailing list, aumenta l'engagement di essi, che saranno quindi più propensi ad acquistare prodotti dall'azienda, o in generale, interagiranno maggiormente con il brand; per questo motivo se una frase ne contiene almeno tre andiamo ad aggiungere un punteggio di 2 al totale.

E3. La frase contiene una virgola.

La presenza di una virgola in una frase, indica che molto probabilmente ci sono 2 periodi, e di conseguenza la probabilità che essa abbia una formulazione confusa aumenta. A questa regola è stato assegnato un peso di 0,5.

E4. La frase contiene più di undici parole.

Come detto prima, tutte le frasi del dataset contengono più di dieci parole; oltre a questo è da considerare che più parole ci sono in una frase più questa diventa complessa. In questo caso il peso assegnato è di 1,5.

Frasi	URL
[...] If you DON'T want to receive these emails, please tick this box. [...] If you DON'T want to receive these brochures, please tick this box. [...]	https://loaf.com/products/humblebum-sofa?size=small&colour=thatch-house-fabric
Uncheck the box if you prefer not to receive email updates.	https://www.bcbg.com/en/desert-flower-strappy-bikini-top/695159793264.html
We'd like to keep you informed via email about product updates, upgrades, special offers and pricing. If you do not wish to be contacted via email, please ensure that the box is not checked.	https://www.gshock.com/watches/master-of-g/gg1000-1a
Tick here if you do not want to receive Promotional Offers by Post Tick here if you do wish to receive Promotional Offers by email and/or SMS Tick here if you would like to receive Offers from selected third parties	https://www.littlewoodsireland.ie/account/register.page?
We'd love to send you emails with offers and new products from New Balance Athletics, Inc. but if you do not wish to receive these updates, please tick this box.	https://www.newbalance.co.uk/nb1/nb1-990v3/US990MC3-BLANK.html?PROD=5912954&designId=5912954
If you click continue below we will be happy to tell you about discounts, money off and new products from Very by e-mail and text, unless you tell us you DO NOT want to hear about them, by ticking here. If you would like to receive offers from carefully selected 3rd parties, please tick here	https://www.very.co.uk/aruba-outdoor-2-seaternbspmodular-sofa-set/1600200267.prd

Tabella 3.1: Frasi etichettate come *trick question* nel dataset rilasciato da Mathur *et al.* [2019]

3.2 Implementazione dell’algoritmo

L’implementazione dell’algoritmo è avvenuta nell’ambiente *Node.js*, cioè un runtime Javascript, che permette di utilizzare questo linguaggio anche al di fuori del contesto web; le librerie utilizzate sono state invece *TypeScript*, ovvero un’estensione di Javascript che lo rende fortemente tipizzato, e *winkJS*, una libreria che implementa alcuni algoritmi di *NLP*.

Nello specifico l’algoritmo prende in input una frase sulla quale vengono effettuate diverse elaborazioni: inizialmente per verificare la regola **E1**, vengono rimossi tutti i segni di punteggiatura tramite un algoritmo di *part of speech tagging*, ovvero una funzione che associa ad ogni token una parte del discorso; così facendo anche se le parole da ricercare sono vicine ad esempio a delle virgolette, vengono considerate correttamente. Successivamente viene effettuato lo *stemming* del testo, ovvero, si cercano di ridurre le forme inflessive e derivate di una parola ad una forma di base comune; in questo modo, ad esempio, parole come *don’t* e *do not* vengono considerate uguali.

Si passa poi alla verifica della regola **E2**, cioè la ricerca di almeno 3 parole chiave definite nella fase precedente; anche questo controllo viene effettuato dopo lo *stemming* del testo originale. Per la regola **E3**, invece, si ricerca una virgola nella frase in input, cioè quella senza alterazioni dovute ad elaborazioni di *NLP*.

Infine, vengono contate le parole che compongono il testo dell’input, per l’euristica **E4**, aggiungendo il punteggio se esse sono più di dieci.

Per il calcolo del punteggio finale, si divide il totale dei punti ottenuti dalla verifica delle euristiche per la somma dei “pesi” delle euristiche, in modo da ottenere un risultato compreso tra zero e uno.

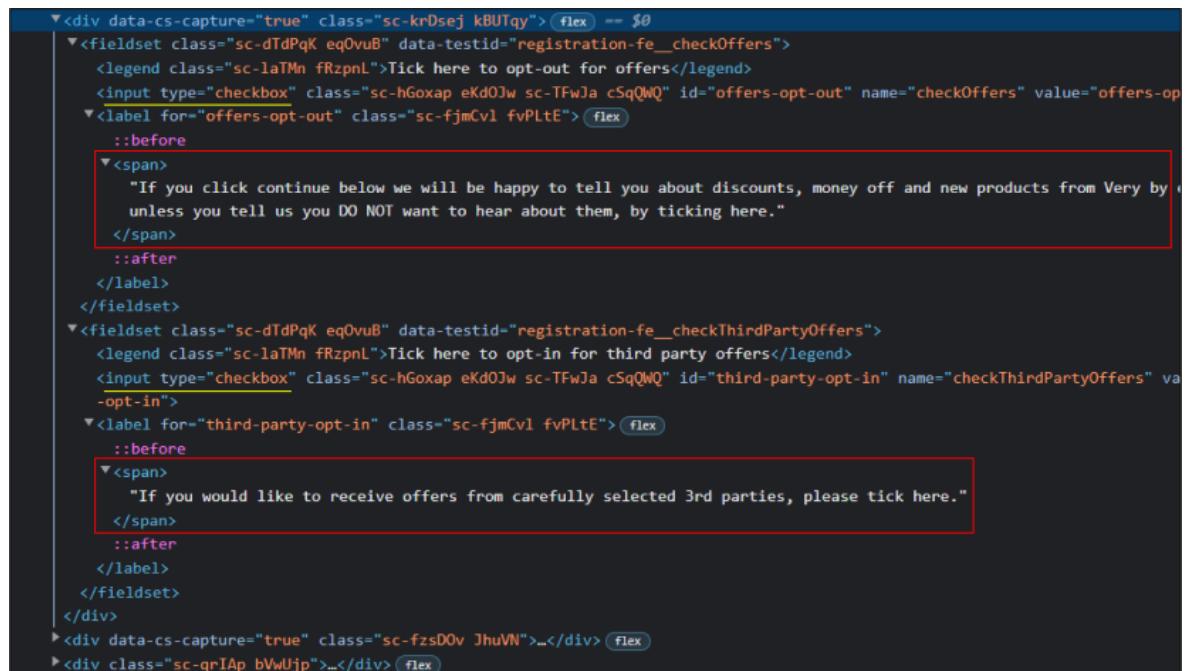
Dal momento che l’algoritmo descritto deve analizzare frasi presenti in siti web, è stato necessario determinare un modo per ricercare queste frasi a partire dal DOM (*Document Object Model*) di una pagina; dato quindi in input un URL di una pagina web, viene effettuata una richiesta HTTP per ottenere il codice HTML, e tramite la libreria *Cheerio*, si attraversa il DOM, alla ricerca di tutti gli elementi che rappresentano delle checkbox, per poi analizzare le frasi nelle immediate vicinanze di essi. Il motivo di questa scelta è che analizzare tutte le frasi presenti nella pagina sarebbe troppo dispendioso in termini di computazione, e di conseguenza richiederebbe troppo tempo; inoltre il DP *trick question* è di solito utilizzato proprio insieme a dei controlli come checkbox, e più raramente radio button. Per ogni checkbox quindi, vengono effettuate delle ricerche in profondità (DFS), considerando come radice dell’albero ogni elemento che si trova sullo stesso livello della checkbox che è stata

trovata; in Figura 3.2 si può notare la corrispondenza tra il DOM e come si presenta la pagina visivamente.

- If you click continue below we will be happy to tell you about discounts, money off and new products from Very by e-mail and text, unless you tell us you DO NOT want to hear about them, by ticking here.

- If you would like to receive offers from carefully selected 3rd parties, please tick here.

(a) Pagina di registrazione



```

<div data-cs-capture="true" class="sc-krDsej kBUTqy" style="flex" -- $0
  <fieldset class="sc-dTdPqK eq0vuB" data-testid="registration-fe__checkOffers">
    <legend class="sc-laTMn fRzpnL">Tick here to opt-out for offers</legend>
    <input type="checkbox" class="sc-hGoxap eKd0Jw sc-TFwJa cSqQWQ" id="offers-opt-out" name="checkOffers" value="offers-opt-out"/>
    <label for="offers-opt-out" class="sc-fjmCvl fvPLtE" style="flex" :before
      <span>
        "If you click continue below we will be happy to tell you about discounts, money off and new products from Very by e-mail and text, unless you tell us you DO NOT want to hear about them, by ticking here."
      </span>
      &:after
    </label>
  </fieldset>
  <fieldset class="sc-dTdPqK eq0vuB" data-testid="registration-fe__checkThirdPartyOffers">
    <legend class="sc-laTMn fRzpnL">Tick here to opt-in for third party offers</legend>
    <input type="checkbox" class="sc-hGoxap eKd0Jw sc-TFwJa cSqQWQ" id="third-party-opt-in" name="checkThirdPartyOffers" value="third-party-opt-in"/>
    <label for="third-party-opt-in" class="sc-fjmCvl fvPLtE" style="flex" :before
      <span>
        "If you would like to receive offers from carefully selected 3rd parties, please tick here."
      </span>
      &:after
    </label>
  </fieldset>
</div>
<div data-cs-capture="true" class="sc-fzsD0v JhuVN">...</div>
<div class="sc-qrIAp bVwUjp">...</div>

```

(b) Struttura del DOM

Figura 3.2: Corrispondenza tra DOM e pagina di registrazione del sito www.very.co.uk

In questo modo anche se una frase si trova ad un livello dell'albero del DOM inferiore rispetto alla checkbox, perché magari è situata all'interno di tag per creare link o decorazioni del testo, viene comunque considerata ed esaminata dall'algoritmo principale che assocerà ad essa un punteggio come descritto precedentemente, e come è possibile notare in Figura 3.3.

(index)	sentence	score
0	' If you click continue below we will be happy to tell you about discounts, money off and new product...' ' If you would like to receive offers from carefully selected 3rd parties, please tick here. ...'	1
1		1

Figura 3.3: Risultati generati dall’algoritmo mostrati nel terminale

Il progetto può essere visualizzato al seguente link su GitHub: <https://github.com/xrenegade100/trick-question-detection>

CAPITOLO 4

Valutazione preliminare dell'approccio

4.1 Metodologia

Per valutare l'efficacia dell'algoritmo, dopo l'implementazione è stato effettuato un processo di testing manuale, tramite il dataset rilasciato da Mathur *et al.* [2019]. Quest'ultimo è strutturato in 7 colonne, contenenti varie informazioni sul dark pattern collezionato, tra cui il pattern vero e proprio, dove applicabile, come frasi ad esempio, la categoria, la pagina web dov'è implementato e un commento generico. Da questa tabella sono state estratte 6 frasi, appartenenti alla categoria *trick question*; la modalità di valutazione non comprende anche la ricerca nella pagina delle frasi, principalmente perché il dataset è stato rilasciato nel 2019, e da quel momento la maggior parte dei siti web, probabilmente dopo essere venuti a conoscenza della loro presenza nella lista, hanno modificato le pagine, rimuovendo i vari dark pattern che erano stati collezionati; inoltre lo scopo principale di questo studio è riuscire a distinguere frasi normali da frasi volutamente confuse, quindi si è scelto di estrarre manualmente le frasi e sottoporle all'analisi dell'algoritmo. La Tabella 3.1 contiene le 6 frasi su cui è stata effettuata la valutazione, riportate integralmente nella prossima sezione. Per ognuna di esse è riportato il punteggio assegnato dall'algoritmo ed eventualmente uno screenshot ottenuto tramite copie cache delle pagine web, per mostrare come erano integrati nelle pagine web.

4.2 Risultati

I risultati ottenuti analizzando le frasi contenute nella Tabella 3.1 sono:

1. "*We'd like to send you weekly emails packed full of our lovely products, clearance goodies and cool stuff to win. You can unsubscribe at any time. If you DON'T want to receive these emails, please tick this box. We'd like to send you the occasional brochure so you can get a regular fix of our lovely product! You can unsubscribe at any time. If you DON'T want to receive these brochures, please tick this box. We'd like to introduce you to other brands you might love who will post you the occasional brochure. You can opt out of this at any time. If you DON'T want to receive these brochures, please tick this box.*". **Punteggio: 1**
2. "*Uncheck the box if you prefer not to receive email updates*". **Punteggio: 0.916**
3. "*We'd like to keep you informed via email about product updates, upgrades, special offers and pricing. If you do not wish to be contacted via email, please ensure that the box is not checked.*".
Punteggio: 1
4. "*Tick here if you do not want to receive Promotional Offers by Post Tick here if you do wish to receive Promotional Offers by email and/or SMS Tick here if you would like to receive Offers from selected third parties*". **Punteggio: 0.916**
5. "*We'd love to send you emails with offers and new products from New Balance Athletics, Inc. but if you do not wish to receive these updates, please tick this box.*". **Punteggio: 1**
6. "*If you click continue below we will be happy to tell you about discounts, money off and new products from Very by e-mail and text, unless you tell us you DO NOT want to hear about them, by ticking here. If you would like to receive offers from carefully selected 3rd parties, please tick here*". **Punteggio: 1** – Questo è l'unico sito nel dataset che attualmente ancora contiene il dark pattern, come si può notare in Figura 4.1. La pagina dove si può visualizzare il modello in azione è <https://www.very.co.uk/account/checkouregister.page?>.

- ✓ Must have between 8 and 40 characters
- ✓ Must have at least one lower case letter
- ✓ Must have at least one upper case letter
- ✓ Must have at least one number

Enter password

[Need help creating a password?](#)

Home address

Find address

If you click continue below we will be happy to tell you about discounts, money off and new products from Very by e-mail and text, unless you tell us you DO NOT want to hear about them, by ticking here.

If you would like to receive offers from carefully selected 3rd parties, please tick here.

By clicking continue you are agreeing to the website [Terms & Conditions](#) and [Privacy Policy](#).

Continue

Figura 4.1: Trick question implementato attualmente su www.very.co.uk

CAPITOLO 5

Conclusioni e sviluppi futuri

L'algoritmo sviluppato con questo studio ha lo scopo di identificare i dark pattern di tipo *trick question* all'interno di pagine web; tuttavia, bisogna considerare che l'approccio utilizzato serve a dar luogo a futuri sviluppi sul tema, sia migliorando eventualmente l'algoritmo con euristiche determinate non solo tramite l'analisi del dataset di Mathur *et al.* [2019], oppure utilizzando altre tecniche di *natural language processing* come *sentiment analysis* o tramite algoritmi di machine learning classificando screenshot di pagine web.

In particolare bisogna tenere conto del fatto che avendo usato un solo dataset per definire le regole per l'identificazione, i risultati potrebbero essere influenzati dal modo in cui esso è stato realizzato; quindi questo studio potrebbe essere esteso prendendo in considerazione dataset diversi e, di conseguenza, ottenere risultati migliori.

Tuttavia l'algoritmo, come illustrato nel Capitolo 4, è efficiente e può essere integrato all'interno di Arkan (<https://www.projectarkan.com>), un progetto realizzato durante il corso di Interazione Uomo-Macchina. Il sito rappresenta un punto di riferimento per gli utenti sul tema dei dark pattern: permette di conoscere il tema tramite esempi e spiegazioni della tassonomia di Gray *et al.* [2018]; inoltre la piattaforma permette di analizzare pagine web alla ricerca dei dark pattern presenti, di varie tipologie, ognuna con un algoritmo dedicato. L'integrazione sarebbe un processo semplice e rapido perché il backend di Arkan è implementato con NodeJS proprio come l'algoritmo di identificazione.

Ringraziamenti

Ci tengo a ringraziare le persone che in un modo o nell'altro hanno fornito un contributo alla realizzazione di questo lavoro.

Innanzitutto, vorrei ringraziare il relatore di questa tesi, il prof. Fabio Palomba per i consigli e gli spunti di riflessione che mi ha dato durante la scrittura di questa tesi, oltre che per la sua infinita disponibilità. Inoltre il suo impegno e la passione nell'insegnare mi hanno permesso di entrare nel mondo dell'ingegneria del software.

Un ringraziamento speciale anche a Giulia Sellitto, seconda relatrice, per l'attenzione che mi ha dedicato durante la scrittura e per l'aiuto che mi ha dato con le sue correzioni.

Un grazie infinito anche alla mia famiglia – i miei genitori, i miei nonni e mio zio – per il supporto che mi hanno dato durante questo percorso e i sacrifici che hanno fatto che mi hanno permesso di arrivare fino a questo punto.

Non posso fare a meno di ringraziare anche Antonio, Giovanni e Manuela, immancabili compagni di avventura, che mi hanno aiutato nei momenti di difficoltà, e con i quali spero di poter raggiungere molti altri successi e traguardi nella vita.

Vorrei ringraziare anche Alfonso e Natale per le infinite risate, ma anche per i discorsi seri e in generale tutti i momenti passati insieme, che mi hanno permesso di comprendere un sacco di concetti e punti di vista differenti.

Un grazie speciale ai ragazzi delle associazioni CoScienze e Iperuranio, che nonostante li abbia conosciuti da poco sono super gentili e simpatici nei miei confronti oltre che sempre disponibili in ogni caso.

Non potrei non menzionare Gaetano, Pasquale ed Antonino, miei fratelli per scelta, oltre che tutti i miei amici di Meta: ognuno a modo proprio si è rivelato speciale, e mi hanno sempre supportato ed aiutato nei momenti di difficoltà.

Per concludere dedico questo traguardo a me stesso, per i sacrifici che ho fatto e faccio ancora, per la mia determinazione e perseveranza che mi hanno fatto arrivare fino a questo momento.

Bibliografia

BÖSCH, C., ERB, B., KARGL, F., KOPP, H. e PFATTHEICHER, S. (2016), «Tales from the dark side: Privacy dark strategies and privacy dark patterns», *Proceedings on Privacy Enhancing Technologies*, vol. 2016 (4), p. 237–254. (Citato a pagina 7)

BRIGNULL, H. (2010), «Dark Patterns», <https://www.darkpatterns.org>. (Citato a pagina 10)

CONTI, G. e SOBIESK, E. (2010), «Malicious interface design: exploiting the user», in «Proceedings of the 19th international conference on World wide web», p. 271–280. (Citato a pagina 7)

DI GERONIMO, L., BRAZ, L., FREGNAN, E., PALOMBA, F. e BACCHELLI, A. (2020), «UI dark patterns and where to find them: a study on mobile applications and user perception», in «Proceedings of the 2020 CHI conference on human factors in computing systems», p. 1–14. (Citato alle pagine 4 e 22)

GRAY, C. M., KOU, Y., BATTLES, B., HOGGATT, J. e TOOMBS, A. L. (2018), «The dark (patterns) side of UX design», in «Proceedings of the 2018 CHI conference on human factors in computing systems», p. 1–14. (Citato alle pagine 8, 10 e 36)

HAMARI, J., KOIVISTO, J. e SARSA, H. (2014), «Does Gamification Work? – A Literature Review of Empirical Studies on Gamification», in «2014 47th Hawaii International Conference on System Sciences», p. 3025–3034. (Citato a pagina 21)

HOEPMAN, J.-H. (2014), «Privacy design strategies», in «IFIP International Information Security Conference», p. 446–459, Springer. (Citato a pagina 8)

MATHUR, A., ACAR, G., FRIEDMAN, M. J., LUCHERINI, E., MAYER, J., CHETTY, M. e NARAYANAN, A. (2019), «Dark patterns at scale: Findings from a crawl of 11K shopping websites», *Proceedings of the ACM on Human-Computer Interaction*, vol. 3 (CSCW), p. 1–32. (Citato alle pagine vi, 5, 27, 29, 33 e 36)