



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΠΛΗΡΟΦΟΡΙΑΚΑ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

**Crowdturfing και E-commerce: Επισκόπηση και μελέτη των
επιπτώσεων σε συστήματα ηλεκτρονικού εμπορίου**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Βασίλειου Χατζηζήση

Επιβλέπων : Ακριβή Βλάχου

Μέλη εξεταστικής επιτροπής: Ακριβή Βλάχου, Δημήτρης Σκούτας, Κωνσταντίνος Μαλιάτσος

Σάμος, [Σεπτέμβριος 2022]

Πρόλογος και ευχαριστίες

Μέσα από αυτή την εργασία, είχα την τύχη και την ευκαιρία μέσα από την τριβή με διάφορα επιστημονικά άρθρα, να δω και να αντιληφθώ διάφορα ζητήματα που προκύπτουν κατά καιρούς με την εξέλιξη της τεχνολογίας που διαρκώς γίνεται πιο προσιτή για τον άνθρωπο και «προσπαθεί» να καλύψει και να ευκολύνει μεγαλύτερη μερίδα αναγκών των ανθρώπων...

Για αυτή μου την ευκαιρία να μπορώ να έχω επαφή με τέτοιου είδους θέματα θα ήθελα να ευχαριστήσω όλους τους ανθρώπους που βρέθηκαν στο διάβα μου και με βοήθησαν να δω την αξία της μόρφωσης (δασκάλους, καθηγητές αλλά και πρότυπα ζωής). Ειδικά όμως θα ήθελα να ευχαριστήσω μέσα από τα βάθη της καρδιάς μου όλους αυτούς (φίλους-οικογένεια) οι οποίοι με μεγάλες ή και μικρές θυσίες με βοήθησαν πρακτικά ώστε να βρίσκω χρόνο μέσα σε δύσκολες συνθήκες να ασχολούμαι με την μόρφωση.

2022

του

ΧΑΤΖΗΖΗΣΗ ΒΑΣΙΛΕΙΟΥ

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Περίληψη

Το crowdurfing αποτελεί απειλή για το σύνολο του Διαδικτύου, όχι μόνο επειδή υπονομεύει την αξιοπιστία των διακινούμενων πληροφοριών, αλλά και επειδή δεν είναι αξιόποιο έγκλημα. Ως εκ τούτου, δύσκολα θα επιτευχθεί η ανακοπή της εν λόγω πρακτικής, καθώς τα τελευταία χρόνια, έχουν εμφανιστεί έρευνες που σημειώνουν ότι το φαινόμενο του crowdsourcing (δηλαδή, η μαζική πρόσληψη ατόμων) με σκοπό το crowdurfing, γίνεται όλο και πιο σύνηθες. Η παρούσα βιβλιογραφική ανασκόπηση επιχειρεί να αναδείξει το ζήτημα του crowdurfing και την επιρροή του στο ηλεκτρονικό εμπόριο (e-commerce). Αρχικά, αναλύονται οι έννοιες των διαδικτυακών αξιολογήσεων και των κριτικών από καταναλωτές, έπειτα ο τρόπος που τις βιώνει ο εκάστοτε καταναλωτής και εν συνεχεία, γίνεται η ανάλυση του φαινομένου του crowdurfing και των τεχνικών αποφυγής του. Τέλος, παρατίθεται η μελέτη περίπτωσης της Amazon σε πλήρη συνάρτηση με τις πρακτικές crowdurfing.

Λέξεις κλειδιά: crowdurfing, e-commerce, astroturfing, crowdsourcing, διαδίκτυο

Πίνακας Περιεχομένων

Περίληψη	5
Εισαγωγή	9
ΚΕΦΑΛΑΙΟ 1 ^ο	13
1.1 Διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών	13
1.2 Λόγοι για να ενεργοποιήσουν οι επιχειρήσεις τις αξιολογήσεις και κριτικές καταναλωτών	14
1.3 Οι ρόλοι που διαδραματίζουν οι επιχειρήσεις στην υποστήριξη αξιολογήσεων και κριτικών καταναλωτών	16
1.4 Διαχείριση αξιολογήσεων και κριτικών καταναλωτών από επιχειρήσεις	17
1.5 Χρηματοδοτικά επιχειρηματικά μοντέλα αξιολόγησης	19
ΚΕΦΑΛΑΙΟ 2 ^ο	20
2.1 Ο τρόπος που βιώνει ο καταναλωτής τις αξιολογήσεις και τις κριτικές στο διαδίκτυο	20
2.1.1 Οφέλη καταναλωτή.....	20
2.2 Χρήση διαδικτυακών αξιολογήσεων και κριτικών από τους καταναλωτές.....	21
2.3 Οι αντιλήψεις των καταναλωτών	22
2.4 Ζητήματα με διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών	25
2.5 Ψεύτικες βαθμολογίες και κριτικές.....	26
2.6 Άλλες παραπλανητικές πρακτικές.....	28
2.6.1 Κριτικές με κίνητρα	29
2.6.2 Συγκεντρωτικές αξιολογήσεις και άλλες αναδυόμενες πρακτικές	30
2.7 Ο ρόλος που διαδραματίζουν οι διαδικτυακές πλατφόρμες.....	32
2.8 Η ακρίβεια των αξιολογήσεων και των κριτικών των καταναλωτών ως ενδείξεις ποιότητας.....	32
2.9 Προκαταλήψεις συμπεριφοράς των καταναλωτών στην παραγωγή και χρήση αξιολογήσεων και κριτικών	34

ΚΕΦΑΛΑΙΟ 3 ^ο	37
3.1 Τι είναι το crowdurfing	37
3.2 Η σκοτεινή πλευρά του crowdsourcing	45
3.3 Ηθικά ζητήματα στη χρήση IS	47
3.3.1 Παραδοσιακή Φιλοσοφική Ηθική.....	47
3.3.2 Επιχειρηματική Ηθική	48
3.3.3 Κωδικοποιημένοι κανόνες	49
3.4 Η σύγκρουση του κυβερνοχώρου και του φυσικού κόσμου.....	50
ΚΕΦΑΛΑΙΟ 4 ^ο	52
4.1 Crowdturfing και Social Media.....	52
4.2 Η διαδικασία ανίχνευσης	54
4.3 Εξόρυξη και δημιουργία προφίλ χρηστών μέσω κοινωνικής δικτύωσης	56
4.4 Μοντελοποίηση διάχυσης πληροφοριών στα μέσα κοινωνικής δικτύωσης	58
4.5 Προσεγγίσεις που βασίζονται στο περιεχόμενο.....	59
4.6 Προσεγγίσεις που βασίζονται στη συμπεριφορά	59
4.7 Προσεγγίσεις που βασίζονται στη διάχυση	60
4.8 Σύνολα δεδομένων	60
4.8 Μελλοντικές κατευθύνσεις	61
ΚΕΦΑΛΑΙΟ 5 ^ο	63
5.1 Επιθέσεις crowdsourced σε συστήματα αξιολόγησης e-commerce.....	63
5.2 Αξιολόγηση ποιότητας κριτικών που έχουν δημιουργηθεί από μηχανήματα....	64
5.3 Άμυνα έναντι κριτικών που δημιουργούνται από μηχανή	68
5.4 Νευρωνικά δίκτυα και ασφάλεια	70
ΚΕΦΑΛΑΙΟ 6 ^ο	72
6.1 Μελέτη περίπτωσης: Amazon.....	72
6.2 Περιγραφή υπόθεσης.....	73

*Διπλωματική Εργασία: Crowdfunding και E-commerce: Επισκόπηση και μελέτη των
επιπτώσεων σε συστήματα ηλεκτρονικού εμπορίου*

6.3 Ανάλυση περίπτωσης	75
6.3.1 Οι πελάτες	75
6.3.2 Εργαζόμενοι.....	76
6.3.3 Μέσο	77
Συζήτηση.....	78
Συμπέρασμα.....	81
Βιβλιογραφία	83

Εισαγωγή

Το Διαδίκτυο δεν είναι πλέον η πηγή αξιόπιστων πληροφοριών που ήταν κάποτε. Σήμερα, η παραπληροφόρηση χρησιμοποιείται ως εργαλείο για να βλάψει τους ανταγωνιστές, να κερδίσει πολιτικές εκστρατείες και να επηρεάσει την κοινή γνώμη. «Συγκρούσεις» μεταξύ λογαριασμών συμβαίνουν καθημερινά στα κοινωνικά δίκτυα και στα διαδικτυακά φόρουμ συζήτησης και η αξιοπιστία πολλών διαδικτυακών πηγών πληροφοριών τίθεται πλέον υπό αμφισβήτηση. Ένα πολύ αποτελεσματικό όπλο για τη διάδοση παραπληροφόρησης είναι η χρήση εκστρατειών crowdurfing, όπου πληρώνονται ομάδες χρηστών για να κάνουν αμφισβητούμενες ή παράνομες ενέργειες στο διαδίκτυο (Lee et al., 2014). Για παράδειγμα, ένας εισβολέας μπορεί να πληρώσει στους εργαζόμενους μικρά ποσά για να γράψουν αρνητικές διαδικτυακές κριτικές για μια ανταγωνιστική επιχείρηση, συχνά κατασκευάζοντας ανύπαρκτους λογαριασμούς. Δεδομένου ότι αυτά είναι γραμμένα από πραγματικούς ανθρώπους, συχνά δεν εντοπίζονται από αυτοματοποιημένα εργαλεία που αναζητούν εισβολείς λογισμικού (Lee et al., 2014).

Ένας πιο πεπειραμένος εισβολέας μπορεί να εφαρμόσει αντίπαλες τεχνικές, ωστόσο οι ταξινομητές μηχανικής μάθησης (machine learning – ML) μπορούν να ανιχνεύσουν αποτελεσματικά αυτές τις επιθέσεις. Ωστόσο, η πρόοδος στη μηχανική μάθηση και στα βαθιά νευρωνικά δίκτυα (deep neural networks - DNN) μπορεί να χρησιμεύσει για να κάνει αυτές τις επιθέσεις πολύ πιο ισχυρές και δύσκολες στην αποφυγή. Συγκεκριμένα, σε περιορισμένα περιβάλλοντα εφαρμογών, τα DNN έχουν φτάσει σε ένα σημείο όπου μπορούν να παράγουν επαρκώς σαφές και σωστό περιεχόμενο που δεν διακρίνεται αποτελεσματικά από αυτό που παράγεται από ανθρώπους. Εστιάζουμε κυρίως στον τομέα των διαδικτυακών κριτικών για προϊόντα και υπηρεσίες ηλεκτρονικού εμπορίου, όπου εκατομμύρια χρήστες ανεβάζουν κριτικές σε ιστότοπους όπως το Yelp, το *Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ*

TripAdvisor και το Amazon. Οι διαδικτυακές κριτικές τείνουν να είναι σύντομες και εστιάζονται σε περιορισμένο εύρος θεμάτων που ορίζονται από τον τομέα της εφαρμογής, π.χ. ποιότητα φαγητού και εξυπηρέτησης σε ένα εστιατόριο.

Τα κακόβουλα φόρουμ crowdsourcing κερδίζουν έδαφος ως πηγές διάδοσης παραπληροφόρησης στο διαδίκτυο, αλλά περιορίζονται από το κόστος πρόσληψης και διαχείρισης ανθρώπινων εργαζομένων. Η έρευνα των Yao et al. (2017) προσδιορίζει μια νέα κατηγορία επιθέσεων που αξιοποιούν επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks - RNN) για να αυτοματοποιήσουν τη δημιουργία ψεύτικων διαδικτυακών κριτικών για προϊόντα και υπηρεσίες. Αυτές οι επιθέσεις όχι μόνο είναι φθηνές και επομένως πιο επεκτάσιμες, αλλά μπορούν να ελέγξουν τον ρυθμό εξόδου περιεχομένου για να εξαλείψουν την υπογραφή που καθιστά εύκολο τον εντοπισμό των crowdsourced καμπανιών. Χρησιμοποιώντας τις κριτικές Yelp ως παράδειγμα πλατφόρμας, διαφαίνεται πώς μια επίθεση δημιουργίας αναθεώρησης και προσαρμογής σε δύο φάσεις μπορεί να παράγει κριτικές που δεν διακρίνονται από στατιστικούς ανιχνευτές τελευταίας τεχνολογίας. Διεξάγεται μια μελέτη χρηστών που βασίζεται σε έρευνες για να τονιστεί ότι αυτές οι αξιολογήσεις όχι μόνο δεν εντοπίζονται από τον άνθρωπο, αλλά λαμβάνουν και υψηλές βαθμολογίες στις μετρήσεις «χρησιμότητας» από τους χρήστες. Τέλος, αναπτύσσονται νέες αυτοματοποιημένες άμυνες ενάντια σε αυτές τις επιθέσεις, αξιοποιώντας τον μετασχηματισμό με απώλειες που εισάγονται από τον κύκλο εκπαίδευσης και παραγωγής RNN. Παράλληλα, λαμβάνουν υπόψη τα αντίμετρα κατά των μηχανισμών, δείχνουν ότι παράγουν μη ελκυστικές ανταλλαγές κόστους-οφέλους για τους εισβολείς και ότι μπορούν να περιοριστούν περαιτέρω από απλούς περιορισμούς που επιβάλλονται από παρόχους διαδικτυακών υπηρεσιών.

Ειδικότερα, η έρευνα των Yao et al. (2014) προσδιορίζει μια κατηγορία επιθέσεων που βασίζονται στη δημιουργία ψεύτικης κριτικής βάσει DNN. Αποδεικνύεται ότι οι γεννήτριες αναθεώρησης που βασίζονται σε DNN είναι πρακτικές και αποτελεσματικές, χρησιμοποιώντας έναν συνδυασμό δημιουργίας αναθεώρησης εκπαιδευμένης σε ML και προσαρμογής βάσει περιβάλλοντος. Χρησιμοποιώντας τις κριτικές εστιατορίων Yelp ως πλατφόρμα-στόχο, δείχνουν εμπειρικά ότι οι συνθετικές κριτικές που δημιουργούνται δεν διακρίνονται ουσιαστικά από τις πραγματικές

κριτικές από ανιχνευτές τελευταίας τεχνολογίας που βασίζονται σε γλωσσικά χαρακτηριστικά. Οι Yao et al. (2014) έδειξαν ότι όχι μόνο αυτές οι ψεύτικες κριτικές μπορούν να αποφύγουν σταθερά τον εντοπισμό από πραγματικούς χρήστες, αλλά παρέχουν το ίδιο επίπεδο «χρησιμότητας» που αντιλαμβάνεται ο χρήστης με τις πραγματικές κριτικές. Στη συνέχεια εξέτασαν πιθανές άμυνες και πρότειναν μια που βασίζεται σε ML-classifier που αξιοποιεί τους εγγενείς υπολογιστικούς περιορισμούς των περισσότερων RNN έναντι του εισβολέα (Yao et al., 2014).

Οι Yao et al. (2014) έδειξαν ότι ο κύκλος επεξεργασίας πραγματικών κριτικών μέσω εκπαίδευσης μοντέλων βάσει RNN και δημιουργίας κειμένου είναι απωλεστικός και η απώλεια που προκύπτει μπορεί να εντοπιστεί συγκρίνοντας την κατανομή σε επίπεδο χαρακτήρων των κριτικών που δημιουργούνται από RNN με εκείνες που γράφτηκαν από πραγματικούς χρήστες. Εξέτασαν επίσης πιθανά αντίμετρα και έδειξαν ότι η αυξανόμενη πολυπλοκότητα του μοντέλου παράγει μειωμένες αποδόσεις στη φοροδιαφυγή, ενώ το κόστος των πόρων αυξάνεται δραματικά. Συνοπτικά, η έρευνά τους παράγει τα εξής βασικά συμπεράσματα (Yao et al., 2014):

(1) Επιδεικνύεται η σκοπιμότητα της αυτοματοποιημένης δημιουργίας κριτικών προϊόντων για ιστότοπους κριτικών στο διαδίκτυο, χρησιμοποιώντας μια προσέγγιση που βασίζεται σε RNN για τη δημιουργία και την προσαρμογή των κριτικών. Η βασική γνώση είναι ότι, ενώ η αυτοματοποιημένη δημιουργία περιεχομένου αυθαίρετου μεγέθους είναι πρόκληση, η δημιουργία συντομότερου κειμένου σε τομείς σταθερών εφαρμογών είναι πιο πρακτική.

(2) Οι συνθετικές αναθεωρήσεις που βασίζονται σε RNN είναι ισχυρές έναντι των στατιστικών τελευταίας τεχνολογίας και των ανιχνευτών που βασίζονται σε ML. Επιπλέον, η μελέτη χρήστη έδειξε ότι δεν διακρίνονται σε μεγάλο βαθμό από τις πραγματικές κριτικές για τους ανθρώπινους αναγνώστες και φαίνεται να παρέχουν παρόμοια επίπεδα «χρησιμότητας» όπως καθορίζονται από τους αναγνώστες.

(3) Προτείνεται μια νέα άμυνα που αξιοποιεί την απώλεια πληροφοριών που είναι εγγενής σε μια εκπαιδευτική διαδικασία RNN για τον εντοπισμό στατιστικά ανιχνεύσιμων παραλλαγών στην κατανομή, σε επίπεδο χαρακτήρων των αναθεωρήσεων που δημιουργούνται από μηχανή. Η άμυνα είναι ισχυρή έναντι των

Διπλωματική Εργασία: Crowdturfing και E-commerce: Επισκόπηση και μελέτη των επιπτώσεων σε συστήματα ηλεκτρονικού εμπορίου

αντίμετρων και η αποφυγή εντοπισμού συνεπάγεται ότι ο εισβολέας πληρώνει ταχέως επιταχυνόμενο κόστος για μειωμένες αποδόσεις.

Πιστεύεται ότι πρόκειται για μια πρακτική νέα επίθεση που μπορεί να έχει σημαντικό αντίκτυπο όχι μόνο σε ιστότοπους αξιολόγησης που δημιουργούνται από χρήστες, όπως το Yelp, αλλά πιθανές επιθέσεις σε πλατφόρμες παραγωγής περιεχομένου, όπως το Twitter και τα διαδικτυακά φόρουμ συζήτησης και ότι αυτά τα αποτελέσματα θα επιστήσουν την προσοχή στο πρόβλημα και θα ενθαρρύνουν περαιτέρω ανάλυση και ανάπτυξη νέων αμυνών.

ΚΕΦΑΛΑΙΟ 1^ο

1.1 Διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών

Οι καταναλωτές επωφελούνται επί μακρόν από την ανταλλαγή απόψεων και εμπειριών σχετικά με την ποιότητα και τη χρηστικότητα των αγαθών και των υπηρεσιών. Με την άνοδο του διαδικτύου, του ηλεκτρονικού εμπορίου και των διαδικτυακών πλατφορμών, οι αξιολογήσεις και οι κριτικές έχουν γίνει ολοένα και πιο διαθέσιμες, διαδεδομένες και με επιρροή. Εκατομμύρια καταναλωτές δημιουργούν και βασίζονται σε αξιολογήσεις και κριτικές που παράγονται από άλλους καταναλωτές για να ενημερώνουν τις αγοραστικές τους αποφάσεις. Οι επιχειρήσεις έχουν αναγνωρίσει τη δύναμή τους να διαμορφώνουν την αγοραστική συμπεριφορά των καταναλωτών και έχουν αναπτύξει συστήματα για τη λήψη, την επιμέλεια και την προβολή διαδικτυακών αξιολογήσεων και κριτικών καταναλωτών, τα οποία έχουν γίνει σήμερα επίσης σημαντικά διαφημιστικά μέσα χωρίς τα οποία πολλές επιχειρήσεις δεν θα μπορούσαν να λειτουργήσουν. Η έννοια των διαδικτυακών κριτικών καταναλωτών ορίζεται ως εξής: «Η γνώμη ή/και η εμπειρία ενός καταναλωτή για ένα προϊόν, μια υπηρεσία ή μια επιχείρηση. Μπορεί κανείς να βρει κριτικές σε εξειδικευμένους ιστότοπους πολλών εμπόρων λιανικής, πλατφορμών λιανικής, πρακτόρων κρατήσεων και συστημάτων αξιόπιστων εμπόρων» (Robson et al., 2013).

Εκτός από τις γραπτές κριτικές των καταναλωτών στο διαδίκτυο, έχει παρατηρηθεί αύξηση στη χρήση κριτικών βίντεο. Οι αξιολογήσεις είναι μη αφηγηματικές εκτιμήσεις αγαθών και υπηρεσιών που προβάλλονται σε μια κλίμακα σημείων ή αστεριών (π.χ. αστερία από το 1 έως το 5). Οι αξιολογήσεις και οι κριτικές εμφανίζονται συχνά το ένα δίπλα στο άλλο. Μερικές φορές, οι επιχειρήσεις συμμετέχουν στη μετάφραση των αφηγηματικών κριτικών σε αξιολογήσεις. Οι αξιολογήσεις παρέχουν λιγότερες πληροφορίες για τα συγκεκριμένα χαρακτηριστικά ενός αγαθού ή μιας υπηρεσίας από τις κριτικές. Διευκολύνουν τους καταναλωτές να συγκρίνουν μεγαλύτερους αριθμούς αγαθών και υπηρεσιών. Μια μελέτη των Archak et al. (2011) προτείνει ότι οι *Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ*

αξιολογήσεις έχουν μικρότερο βάρος στη διαδικασία λήψης αποφάσεων των καταναλωτών από τις κριτικές. Ωστόσο, μπορεί να είναι δύσκολο να διαχωριστεί η συνολική επίδραση των αξιολογήσεων και των κριτικών, καθώς συχνά παρουσιάζονται μαζί. Πρόσφατα, αντίστροφες αξιολογήσεις και κριτικές (μέρος συστημάτων αμφίπλευρης ανάδρασης) αναπτύχθηκαν και εφαρμόστηκαν από επιχειρήσεις (π.χ. Uber και Airbnb) για να βαθμολογήσουν τους πελάτες τους.

Σε ορισμένα συστήματα, αυτές οι αξιολογήσεις και οι κριτικές δημιουργούν εμπιστοσύνη στον πωλητή για να προσφέρει το αγαθό ή την υπηρεσία του (π.χ. να νοικιάσει ένας ιδιοκτήτης ένα διαμέρισμα ή ένα σπίτι). Οι αξιολογήσεις και οι κριτικές των καταναλωτών διαφέρουν από τις κριτικές που παράγονται από ειδικούς (δηλαδή άτομα, ομάδες ή ιδρύματα που κατέχουν, ως αποτέλεσμα της εμπειρίας, της μελέτης ή της εκπαίδευσής τους, γνώση ενός συγκεκριμένου θέματος, η οποία γνώση είναι ανώτερη από ό,τι αποκτούν γενικά τα συνηθισμένα άτομα). Αυτές οι αξιολογήσεις έχουν γενικά μεγαλύτερη βαρύτητα στη διαδικασία λήψης αποφάσεων των καταναλωτών. Διαφέρουν επίσης από τις εγκρίσεις που δημιουργούνται συνήθως από άτομα με εξουσία, όπως διασημότητες και άτομα που επηρεάζουν τα μέσα κοινωνικής δικτύωσης. Οι επηρεαστές (influencers) των μέσων κοινωνικής δικτύωσης μπορούν να είναι ειδικοί στη γυμναστική, beauty bloggers, fashionistas και ειδικοί στα ταξίδια. Οι εγκρίσεις τους έχουν γενικά μεγαλύτερη βαρύτητα στη διαδικασία λήψης αποφάσεων των καταναλωτών από ό,τι οι εγκρίσεις από άλλους ομοτίμους καταναλωτές, λόγω της εμπειρίας, της αυθεντικότητας ή της εξουσίας τους. Συχνά δίνονται κίνητρα για τις εγκρίσεις που γίνονται από ειδικούς, διασημότητες και άτομα που επηρεάζουν τα μέσα κοινωνικής δικτύωσης (Robson et al., 2013). Η παρούσα ενότητα περιγράφει ορισμένα σημαντικά χαρακτηριστικά και τη δυναμική της αγοράς σχετικά με τις αξιολογήσεις και τις κριτικές των καταναλωτών στο διαδίκτυο.

1.2 Λόγοι για να ενεργοποιήσουν οι επιχειρήσεις τις αξιολογήσεις και κριτικές καταναλωτών

Οι επιχειρήσεις ενθαρρύνουν ενεργά τις αξιολογήσεις και τις κριτικές των καταναλωτών στο διαδίκτυο για διάφορους λόγους. Οι διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών επιτρέπουν στους πιθανούς αγοραστές να παρακολουθούν την προηγούμενη απόδοση ενός πωλητή και να λαμβάνουν ταχύτερες και πιο

ενημερωμένες αποφάσεις. Επίσης, «κατευθύνουν» τις αρχικές πωλήσεις, την ικανοποίηση των πελατών, τις επαναλαμβανόμενες πωλήσεις και τελικά το κέρδος, καθώς και την αξία των μετόχων (De Langhe et al., 2016).

Σύμφωνα με μια μελέτη των ΗΠΑ του 2013, το 96% των εμπόρων λιανικής κατατάσσουν τις αξιολογήσεις και τις κριτικές των καταναλωτών ως αποτελεσματικό εργαλείο για την αύξηση της μετατροπής. Μπορούν να αυξήσουν την κατάταξη μιας επιχείρησης στις πλατφόρμες αναζήτησης. Αυτό ισχύει ιδιαίτερα όσον αφορά τις διαδικτυακές κριτικές ή τις εγκρίσεις που γίνονται από άτομα που επηρεάζουν τα μέσα κοινωνικής δικτύωσης, οι οποίοι συχνά συνεργάζονται με επώνυμα προϊόντα και προωθούν προϊόντα στο κοινό ή στους «ακόλουθους» τους στα μέσα κοινωνικής δικτύωσης. Ειδικά οι μικρές και μεσαίες επιχειρήσεις μπορεί να διαπιστώσουν ότι οι αξιολογήσεις και οι κριτικές σε διαδικτυακές πλατφόρμες έχουν σημαντικό αντίκτυπο στις πωλήσεις. Οι επιχειρήσεις μπορούν να χρησιμοποιούν αξιολογήσεις και κριτικές καταναλωτών ως μέρος της στρατηγικής διαφήμισης και μάρκετινγκ. Μπορούν να παρέχουν στους καταναλωτές πληροφορίες σχετικά με την απόδοση του προϊόντος σε «περιβάλλον» λέξεων και να παρέχουν στους καταναλωτές πληροφορίες που σχετίζονται με τη χρήση του προϊόντος ή της υπηρεσίας για την οποία προορίζονται. Ως εκ τούτου, μπορούν να βοηθήσουν στην ενστάλαξη εμπιστοσύνης σε πωλητές που μπορεί να μην είναι γνωστοί στους πιθανούς αγοραστές. Τα υψηλά επίπεδα εμπιστοσύνης των καταναλωτών μπορούν να ενισχύσουν την επιχειρηματική φήμη, γεγονός που επιτρέπει στην επιχείρηση να προσελκύει περισσότερους αγοραστές και να χρεώνει υψηλότερες τιμές (De Langhe et al., 2016).

Πράγματι, οι διαδικτυακές πλατφόρμες έχουν χρησιμοποιήσει συστήματα αξιολόγησης καταναλωτών ως εκ των υστέρων μηχανισμό οικοδόμησης εμπιστοσύνης. Η ακαδημαϊκή βιβλιογραφία σε διαδικτυακές πλατφόρμες όπως το eBay και το Amazon έχει σημειώσει ότι η φήμη ενός πωλητή επηρεάζει τη ζήτηση. Στο πλαίσιο των αγορών ομοτίμων πλατφορμών, τα ευρήματα από μια έρευνα του ΟΟΣΑ του 2017 υποδηλώνουν ότι η χρήση συστημάτων φήμης είναι ένας από τους πιο σημαντικούς μηχανισμούς οικοδόμησης εμπιστοσύνης και βασικό εργαλείο για την ενίσχυση της δέσμευσης των καταναλωτών σε αγορές ομοτίμων πλατφορμών. Επιπλέον, τόσο οι θετικές όσο και οι αρνητικές κριτικές έχουν ουσιαστικό αντίκτυπο στη φήμη και τις

πωλήσεις μιας επιχείρησης. Οι αρνητικές κριτικές θα μπορούσαν να έχουν σημαντικές αρνητικές επιπτώσεις σε μια επιχείρηση: οι καταναλωτές γενικά προσέχουν τις αρνητικές κριτικές περισσότερο από τις θετικές κριτικές. Ωστόσο, υπάρχουν ορισμένες ενδείξεις ότι οι καταναλωτές είναι πιο πιθανό να αγοράσουν προϊόντα που έχουν ένα μείγμα κυρίως καλών και μερικών κακών κριτικών αντί για προϊόντα που έχουν μόνο καλές κριτικές, επειδή οι καταναλωτές μπορεί να είναι καχύποπτοι εάν υπάρχουν μόνο καλές κριτικές. Όταν υπάρχουν αρνητικές κριτικές, η συνολική πηγή των αξιολογήσεων φαίνεται πιο γνήσια (Wang et al., 2020).

1.3 Οι ρόλοι που διαδραματίζουν οι επιχειρήσεις στην υποστήριξη αξιολογήσεων και κριτικών καταναλωτών

Υπάρχουν πολλοί διαφορετικοί ρόλοι που μπορούν να παίξουν οι αξιολογήσεις και οι κριτικές καταναλωτών στο διαδίκτυο στο πλαίσιο των επιχειρήσεων. Ωστόσο, η διάκριση μεταξύ αυτών είναι χρήσιμη για την κατανόηση των διαφορετικών ενδιαφερόντων και κινήτρων, καθώς και της δυναμικής στην αγορά. Μπορούν να αναγνωριστούν τουλάχιστον τέσσερις διαφορετικοί ρόλοι (Wang et al., 2020):

-Τύπος 1: Επιχειρήσεις που πωλούν τα δικά τους προϊόντα ή/και υπηρεσίες διαδικτυακά. Οι ιστότοποι τους περιέχουν αξιολογήσεις καταναλωτών και κριτικές για τα προϊόντα και τις υπηρεσίες τους.

-Τύπος 2: Ιστότοποι σύγκρισης τιμών, ιστότοποι κρατήσεων, ηλεκτρονικές μηχανές αναζήτησης (όπως η Google) και ιστότοποι από ενώσεις καταναλωτών που διευκολύνουν τις αξιολογήσεις καταναλωτών και τις κριτικές για τα προϊόντα ή/και τις υπηρεσίες που συγκρίνονται, βαθμολογούνται ή ελέγχονται. Δεν πραγματοποιούν κατ' ανάγκη απευθείας πωλήσεις σε καταναλωτές.

-Τύπος 3: Επιχειρήσεις που παρέχουν υποδομή για αξιολογήσεις καταναλωτών και αξιολογήσεις σε άλλες επιχειρήσεις, όπως η ενσωμάτωση αξιολογήσεων καταναλωτών και προγραμμάτων κριτικών σε ιστότοπους πωλητών. Τέτοιες επιχειρήσεις μπορούν να προσφέρουν υπηρεσίες για τη διαχείριση αξιολογήσεων ή να παρέχουν πρόσβαση σε

βάσεις δεδομένων καταναλωτών που είναι πρόθυμοι να αξιολογήσουν αγαθά και υπηρεσίες, με σκοπό την προσέλκυση και την παροχή κινήτρων στους νέους καταναλωτές για αξιολόγηση. Αυτός ο τύπος επιχειρήσεων δεν πωλεί απευθείας αγαθά ή/και υπηρεσίες σε καταναλωτές.

-Τύπος 4: Διαδικτυακές πλατφόρμες, όπως οι υπηρεσίες μέσω κοινωνικής δικτύωσης, που έχουν την ευελιξία να επιτρέπουν στους καταναλωτές να αφήνουν αξιολογήσεις και κριτικές για προϊόντα άλλων επιχειρήσεων. Αποτελούν επίσης περιβάλλοντα για τις επιχειρήσεις τύπου 3.

Επιπλέον, ένας νέος τύπος αναθεώρησης έχει εμφανιστεί την τελευταία δεκαετία: επιχειρήσεις που παρέχουν υπηρεσίες διαχείρισης φήμης. Όπως είναι λογικό, ορισμένες επιχειρήσεις επιδιώκουν να προστατεύσουν την εικόνα και την επωνυμία τους από ζημιές που προκαλούνται από ανταγωνιστές που θα μπορούσαν να δημιουργήσουν ή να δώσουν κίνητρο για ψεύτικες αρνητικές κριτικές. Ως εκ τούτου, συμμετέχουν σε αξιολογήσεις, άμεσα ή μέσω εταιρειών διαχείρισης φήμης, για να αποτρέψουν ή να τερματίσουν ορισμένες συμπεριφορές καταναλωτών. Τέτοιες συμπεριφορές περιλαμβάνουν, για παράδειγμα:

-Αποτροπή των καταναλωτών από τη χρήση αρνητικών αξιολογήσεων και κριτικών για να απαιτήσουν δωρεάν προϊόντα, υπηρεσίες ή άλλα οφέλη από την επιχείρηση.

-Αποτροπή ή τερματισμός των επιθετικών αλληλεπιδράσεων των καταναλωτών ή της χρήσης υβριστικής γλώσσας στις κριτικές τους.

1.4 Διαχείριση αξιολογήσεων και κριτικών καταναλωτών από επιχειρήσεις

Η διαδικασία μέσα από την οποία οι επιχειρήσεις διαχειρίζονται τις αξιολογήσεις περιλαμβάνει τα ακόλουθα βήματα (Wang et al., 2020):

-Επαλήθευση: Οι υποβληθείσες αξιολογήσεις ή κριτικές θα μπορούσαν να δημοσιευτούν απευθείας, αλλά συχνά θα επαληθεύονται από τις επιχειρήσεις εκ των προτέρων.

-Εποπτεία: Ως δεύτερο βήμα, το περιεχόμενο μιας κριτικής ή αξιολόγησης μπορεί να ελεγχθεί. Οι επιχειρήσεις θα έχουν διαφορετικά κριτήρια για την εποπτεία. Αυτό μπορεί να οδηγήσει στην αφαίρεση/απόρριψη ή στην αποδοχή μιας αξιολόγησης.

-Επίλυση παραπόνων: Ορισμένες επιχειρήσεις επικοινωνούν με καταναλωτές που έχουν γράψει αρνητικές αξιολογήσεις ή κριτικές σε μια προσπάθεια να λύσουν το πρόβλημα του καταναλωτή, πιθανώς με την ελπίδα να δημιουργήσουν μια θετική αξιολόγηση ή κριτική.

Τα βήματα αυτής της διαδικασίας μπορούν να διεξαχθούν χειροκίνητα ή με αυτοματοποιημένο τρόπο μέσω αλγορίθμων. Μερικές φορές, η διαδικασία είναι ένας συνδυασμός και των δύο. Όλο και περισσότερο, οι επιχειρήσεις συνάπτουν συμβάσεις με τρίτα μέρη για τη διαχείριση της αξιολόγησης των καταναλωτών και της διαδικασίας ελέγχου για αυτούς (τύπος 3). Τα δύο κύρια συστήματα που λειτουργούν για διαδικτυακές αξιολογήσεις και αξιολογήσεις καταναλωτών μπορούν να χαρακτηριστούν κυρίως ως «ανοικτά» ή «κλειστά» συστήματα. Στο πρώτο σύστημα, οποιοσδήποτε μπορεί να δημοσιεύσει αξιολογήσεις και κριτικές, ενώ στο δεύτερο σύστημα αξιολογήσεις και κριτικές μπορούν να αναρτηθούν μόνο από επιβεβαιωμένους αγοραστές. Στην πράξη, υπάρχουν επίσης υβριδικά συστήματα στα οποία το ένα σύστημα ενσωματώνει χαρακτηριστικά του άλλου (Robson et al., 2013).

Για παράδειγμα, ένα ανοιχτό σύστημα θα μπορούσε να εμφανίζει αξιολογήσεις και κριτικές τόσο από επιβεβαιωμένους όσο και από μη επιβεβαιωμένους αγοραστές. Το 2009, η Ομοσπονδιακή Επιτροπή Εμπορίου των Η.Π.Α. (Federal Trade Commission - FTC) κατέστησε σαφές ότι οι διαδικτυακές αξιολογήσεις και οι κριτικές που παρακινήθηκαν από μια επιχείρηση καλύπτονταν από τους Οδηγούς Επικύρωσης της FTC των ΗΠΑ. Επιπλέον, η υπηρεσία διευκρίνισε ότι θα θεωρούσε τις αξιολογήσεις των καταναλωτών και τις κριτικές που γράφτηκαν σε αντάλλαγμα για δωρεάν αγαθά και υπηρεσίες ως διαφημίσεις που υπόκεινται στην επιβολή της FTC. Μέχρι σήμερα, πολλές δικαιοδοσίες έχουν ακολουθήσει παρόμοια προσέγγιση (Robson et al., 2013).

1.5 Χρηματοδοτικά επιχειρηματικά μοντέλα αξιολόγησης

Τα συστήματα αξιολόγησης καταναλωτών χρηματοδοτούνται με διάφορους τρόπους, μεταξύ των οποίων (Wang et al., 2020):

- Προϋπολογισμοί μάρκετινγκ (επιχειρήσεις τύπου 1).
- Διαφημίσεις στον ιστότοπο (επιχειρήσεις τύπου 1 και τύπου 4).
- Χρέωση κλικ προς αριθμό εμφανίσεων (επιχειρήσεις τύπου 2, 3 και 4).
- Πληρωμές που λαμβάνονται από συνδρομές ή πωλήσεις υποδομής αξιολόγησης (επιχειρήσεις τύπου 3 και 4).

Θα πρέπει να σημειωθεί ότι οι επιχειρήσεις ενδέχεται να βιώσουν ένταση ή ακόμα και σύγκρουση μεταξύ των βραχυπρόθεσμων συμφερόντων τους για τη δημιουργία περισσότερων πωλήσεων χειραγωγώντας αξιολογήσεις και των μακροπρόθεσμων συμφερόντων δημιουργίας εμπιστοσύνης, υποστηρίζοντας την ποιότητα και την αξιοπιστία των διαδικτυακών αξιολογήσεων και κριτικών καταναλωτών. Η ισορροπία μεταξύ των δύο συμφερόντων μπορεί να επηρεαστεί από το επίπεδο ηθικής επίγνωσης των εμπλεκόμενων επιχειρήσεων, ειδικά επειδή η ικανότητα του καταναλωτή να εντοπίζει ψεύτικες και παραπλανητικές αξιολογήσεις και κριτικές μπορεί να είναι περιορισμένη ή ανύπαρκτη (Robson et al., 2013).

ΚΕΦΑΛΑΙΟ 2^ο

2.1 Ο τρόπος που βιώνει ο καταναλωτής τις αξιολογήσεις και τις κριτικές στο διαδίκτυο

Αυτή η ενότητα περιγράφει τον τρόπο με τον οποίο οι καταναλωτές που παρέχουν αξιολογήσεις και κριτικές στο διαδίκτυο βιώνουν και σχετίζονται με αυτές. Καθορίζει επίσης τα οφέλη αυτών των διαδικτυακών κριτικών καταναλωτών για άλλους καταναλωτές.

2.1.1 Οφέλη καταναλωτή

Οι διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών παρέχουν τόσο στους άλλους καταναλωτές όσο και στις επιχειρήσεις μια σειρά από πλεονεκτήματα, τα οποία επιτρέπουν (Robson et al., 2013):

- Καλύτερες και ταχύτερες αποφάσεις αγορών.
- Ενισχυμένο ανταγωνισμό μεταξύ των επιχειρήσεων.
- Περιορισμό της αναζήτησης των καταναλωτών.
- Πρόσβαση των καταναλωτών σε ένα ευρύτερο φάσμα προϊόντων.
- Ανταλλαγή απόψεων και εμπειριών μεταξύ συνομηλίκων.

Επιπλέον, οι διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών επιτρέπουν στους καταναλωτές να μοιράζονται απόψεις και εμπειρίες με τεράστιο αριθμό άλλων καταναλωτών, χωρίς άμεση παρέμβαση από διαφημιστές και επιχειρήσεις. Παρέχουν στους καταναλωτές πρόσθετα μέσα αξιολόγησης της ποιότητας των προϊόντων στο πλαίσιο της διαδικασίας αγοράς, ακόμη και όταν μια αγορά γίνεται τελικά εκτός διαδικτύου. Επίσης, οι διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών βοηθούν

στην επίλυση των ασυμμετριών πληροφοριών μεταξύ επιχειρήσεων και καταναλωτών. Μπορούν να δημιουργήσουν έναν βρόχο ανατροφοδότησης για τις επιχειρήσεις, ο οποίος μπορεί να ενημερώσει τη συνεχή βελτίωση των προϊόντων τους. Επιπλέον, οι διαδικτυακές αξιολογήσεις και κριτικές των καταναλωτών διευκολύνουν τη σύγκριση προϊόντων (υπηρεσιών), γεγονός που μπορεί να οδηγήσει σε μεγαλύτερο ανταγωνισμό, χαμηλότερες τιμές, καλύτερη ποιότητα και βελτιωμένη εμπειρία καταναλωτή. Ειδικότερα, οι διαδικτυακές αξιολογήσεις και κριτικές των καταναλωτών διαδραματίζουν σημαντικό ρόλο σε σχέση με την ποιότητα και τη χρήση των προϊόντων που οι καταναλωτές δεν μπορούν να εξετάσουν και να αξιολογήσουν καλά λόγω ασυμμετριών πληροφοριών. Εκτός από τα αγαθά εμπειρίας και αξιοπιστίας, τα παραδείγματα περιλαμβάνουν υπηρεσίες ειδικών, όπως ιατρικές διαδικασίες ή επισκευές αυτοκινήτων, των οποίων η αποτελεσματικότητα μπορεί να προσδιοριστεί μόνο με την πάροδο του χρόνου (Kumar & Purbey, 2018).

2.2 Χρήση διαδικτυακών αξιολογήσεων και κριτικών από τους καταναλωτές

Τα διαθέσιμα δεδομένα υποδηλώνουν ότι οι καταναλωτές ελέγχουν και διαβάζουν διαδικτυακές αξιολογήσεις και κριτικές πριν λάβουν αποφάσεις συναλλαγών. Μια έρευνα του ΟΟΣΑ του 2017 αποκάλυψε ότι περίπου το 70% των καταναλωτών θεωρούν τις αξιολογήσεις και τις κριτικές ως κρίσιμες ή πολύ σημαντικές για τις συναλλαγές τους, με μικρές μόνο διαφορές εμφανείς ανά χώρα. Η Google διαπίστωσε ότι οι αναζητήσεις από κινητές συσκευές για αξιολογήσεις προϊόντων και κριτικές αυξήθηκαν κατά 35% από το 2015 έως το 2017. Στις Ηνωμένες Πολιτείες, μια μελέτη του 2013 διαπίστωσε ότι το 63% των καταναλωτών είναι πιο πιθανό να αγοράσουν από έναν ιστότοπο που περιέχει αξιολογήσεις προϊόντων και κριτικές. Μια έρευνα στις ΗΠΑ που πραγματοποιήθηκε το 2016 έδειξε περαιτέρω ότι το 82% των Αμερικανών λένε ότι μερικές φορές συμβουλευονται διαδικτυακές αξιολογήσεις και κριτικές όταν αγοράζουν κάτι για πρώτη φορά, συμπεριλαμβανομένου του 40% που λένε ότι το κάνουν πάντα ή σχεδόν πάντα. Μια έρευνα καταναλωτών που διεξήχθη το 2018 στην Ιαπωνία έδειξε ότι οι καταναλωτές στη χώρα πάντα (34%) ή συχνά (50%) ελέγχουν τις

αξιολογήσεις και τις κριτικές στο διαδίκτυο πριν κάνουν μια αγορά. Στο Ηνωμένο Βασίλειο, έρευνα που διεξήχθη το 2015 έδειξε ότι η ετήσια καταναλωτική δαπάνη της χώρας των 23 δισεκατομμυρίων GBP επηρεάζεται από διαδικτυακές αξιολογήσεις και κριτικές και ότι το 54% των καταναλωτών χρησιμοποιεί αξιολογήσεις και κριτικές. Μια μελέτη που αναπτύχθηκε το 2018 αποκάλυψε περαιτέρω ότι το 86% των καταναλωτών στο Ηνωμένο Βασίλειο διαβάζουν κριτικές για τοπικές επιχειρήσεις (Wang et al., 2020).

Οι αξιολογήσεις και οι κριτικές των καταναλωτών στο διαδίκτυο έχουν επίσης σημαντικές επιπτώσεις στη συμπεριφορά των καταναλωτών εκτός σύνδεσης. Η Pew Research διαπίστωσε ότι, το 2016, το 45% των καταναλωτών ελέγχει τις κριτικές πριν αγοράσει σε ένα φυσικό κατάστημα. Ωστόσο, υπάρχουν επίσης ορισμένα στοιχεία που υποδηλώνουν ότι οι καταναλωτές δεν βασίζονται απαραίτητα σε διαδικτυακές αξιολογήσεις και κριτικές. Για παράδειγμα, μια έρευνα καταναλωτών του 2017 στα κράτη μέλη της ΕΕ αποκάλυψε ότι μόνο το 22% των καταναλωτών στις αγορές ομότιμων πλατφορμών χρησιμοποιεί πάντα ένα σύστημα αξιολόγησης πριν από μια συναλλαγή, ενώ το 27% το χρησιμοποιεί συχνά. Από την άλλη πλευρά, η πλειοψηφία των καταναλωτών δεν εξετάζει τακτικά τους ομοτίμους παρόχους. Μόνο το 26% των καταναλωτών μερικές φορές επαληθεύει τους ομοτίμους παρόχους πριν από μια συναλλαγή, ενώ ένα άλλο 32% δεν το κάνει ποτέ ή σπάνια (Wang et al., 2020).

2.3 Οι αντιλήψεις των καταναλωτών

Οι καταναλωτές φαίνεται να θεωρούν γενικά τις διαδικτυακές αξιολογήσεις και κριτικές ως σημαντικές και αξιόπιστες πηγές πληροφοριών. Για παράδειγμα, μια μελέτη του 2012 στις Ηνωμένες Πολιτείες αποκάλυψε ότι το 65% των δυνητικών καταναλωτών επέλεξαν ένα προϊόν που δεν ήταν στο αρχικό τους σύνολο εξέτασης μετά από έλεγχο των διαδικτυακών κριτικών καταναλωτών. Μια μελέτη του 2015 δείχνει ότι οι καταναλωτές είναι πιθανό να ξοδέψουν 31% περισσότερα σε επιχειρήσεις με εξαιρετικές κριτικές και το 86% των ανθρώπων θα διστάσουν να αγοράσουν από μια επιχείρηση που έχει αρνητικές διαδικτυακές κριτικές. Η μελέτη Ipsos Global Trends του 2017 διαπίστωσε ότι το 70% των καταναλωτών σε 23 χώρες συμφωνεί με *Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ*

τη δήλωση: «Αυτό που διαβάζω για τις καλές ή κακές εμπειρίες άλλων ανθρώπων επηρεάζει τις εταιρείες ή τις επωνυμίες που επιλέγω» (Roushneh et al., 2022).

Οι αξιολογήσεις και οι κριτικές φαίνεται να διευκολύνουν την εμπιστοσύνη στους πωλητές/παρόχους, και σε μεγάλο βαθμό οι καταναλωτές φαίνεται να εμπιστεύονται τις αξιολογήσεις και τις κριτικές. Σε μια έρευνα που διεξήχθη το 2013 στο Ηνωμένο Βασίλειο, 8 στους 10 καταναλωτές δήλωσαν ότι εμπιστεύονται τις διαδικτυακές κριτικές όσο και τις προσωπικές συστάσεις. Μια μελέτη του 2014 στη Γερμανία δείχνει ότι το 86% των καταναλωτών θεωρεί τις διαδικτυακές κριτικές «αξιόπιστες» ή «πολύ αξιόπιστες». Μια μελέτη των ΗΠΑ του 2016 διαπίστωσε ότι το 51% των καταναλωτών που διαβάζουν online κριτικές λένε ότι γενικά δίνουν μια ακριβή εικόνα της πραγματικής ποιότητας του προϊόντος, ενώ το 65% των καταναλωτών που διαβάζουν πάντα ή σχεδόν πάντα διαδικτυακές αξιολογήσεις και κριτικές πιστεύουν ότι είναι γενικά ακριβείς. Μια άλλη έρευνα στο Ηνωμένο Βασίλειο το 2013 έδειξε ότι το 64% των καταναλωτών βρίσκουν τις απόψεις άλλων καταναλωτών τις πιο αξιόπιστες πηγές πληροφοριών (Roushneh et al., 2022).

Ωστόσο, ο βαθμός στον οποίο οι διαδικτυακές αξιολογήσεις και κριτικές αποτελούν μια ενιαία ή κυρίαρχη πηγή πληροφοριών για τους καταναλωτές παραμένει ασαφής. Η μελέτη του ΟΟΣΑ του 2017 για πλατφόρμες ομοτίμων υποδηλώνει ότι ενώ το 73% των ερωτηθέντων καταναλωτών θεώρησαν την ικανότητα να ελέγχουν τις διαδικτυακές αξιολογήσεις καταναλωτών ως αξιολογήσεις ως «κρίσιμη» ή «πολύ σημαντική» για την προθυμία τους να συμμετάσχουν σε μια συναλλαγή, μεγαλύτερα ποσοστά των ερωτηθέντων έδωσαν προτεραιότητα στην ασφαλή πληρωμή μεθόδους (81%) και την ασφάλεια και το απόρρητο των προσωπικών τους δεδομένων (78%). Η μελέτη δείχνει επίσης ότι, στις αγορές ομοτίμων πλατφορμών, οι αξιολογήσεις δεν είναι πάντα κρίσιμες για την απόφαση για το αν θα προχωρήσει κάποιος σε μια συναλλαγή, αν και θεωρούνται σημαντικές πηγές πληροφοριών (Roushneh et al., 2022).

Άλλες έρευνες δείχνουν ότι οι διαδικτυακές κριτικές μπορεί να υπερισχύουν άλλων δεικτών ποιότητας, όπως η τιμή και το εμπορικό σήμα, υπονοώντας ότι οι καταναλωτές μπορεί να δίνουν μεγάλη σημασία στις διαδικτυακές κριτικές όταν λαμβάνουν αποφάσεις αγοράς (De Langhe et al., 2016). Μια έρευνα του 2018 δείχνει ότι ενώ η

εμπιστοσύνη σε μια επωνυμία παίζει ισχυρό ρόλο στον καθορισμό του πού ψωνίζουν οι καταναλωτές, βασίζονται σε μεγάλο βαθμό στις απόψεις άλλων ανθρώπων για να αποφασίσουν τι θα αγοράσουν. Υπάρχουν επίσης στοιχεία ότι οι καταναλωτές έχουν μικτή άποψη σχετικά με την αξιοπιστία των διαδικτυακών αξιολογήσεων και κριτικών. Μια μελέτη του 2017 από την Ευρωπαϊκή Επιτροπή (European Commission - EC) δείχνει ότι οι καταναλωτές που χρησιμοποιούν πλατφόρμες από ομοτίμους δεν χρησιμοποιούν συστηματικά συστήματα αξιολόγησης από ομοτίμους και δεν τους εμπιστεύονται πάντα. Η μελέτη της EC δείχνει ότι η περιορισμένη εμπιστοσύνη των καταναλωτών στις πλατφόρμες ομοτίμων μπορεί να εξηγηθεί από διάφορους παράγοντες, όπως η υποκειμενικότητα των κριτικών, η αβεβαιότητα σχετικά με την αντιπροσωπευτικότητα και ο φόβος για ψεύτικες κριτικές. Η μελέτη υποδηλώνει επίσης ότι η εμπιστοσύνη των καταναλωτών στις ομότιμες πλατφόρμες φαίνεται να είναι συμφραζόμενη και οι καταναλωτές χρησιμοποιούν στην πραγματικότητα ορισμένα στοιχεία για να αξιολογήσουν τη συνολική αξιοπιστία μιας πλατφόρμας και των ομοτίμων παρόχων της (Poushneh et al., 2022).

Παρόλο που η πλειονότητα των καταναλωτών εκτιμά τις κριτικές και τις αξιολογήσεις όταν εξετάζει εάν θα χρησιμοποιήσει έναν πωλητή/πάροχο, οι καταναλωτές αναγνωρίζουν ότι οι κριτικές και οι αξιολογήσεις δεν μπορούν πάντα να είναι αξιόπιστες, κυρίως λόγω ανέντιμων ή ψεύτικων κριτικών. Μια μελέτη αγοράς του 2017 στη Νορβηγία ρίχνει φως στην εμπιστοσύνη των καταναλωτών στις διαδικτυακές αξιολογήσεις και κριτικές. Η μελέτη προτείνει ότι οι μεροληπτικές αξιολογήσεις καταναλωτών και οι κριτικές για προϊόντα σε αγορές ομοτίμων πλατφορμών όπου οι καταναλωτές τείνουν να αποφεύγουν να δίνουν αρνητικές αξιολογήσεις από φόβο μήπως βλάψουν την καθιερωμένη προσωπική σχέση με τους πωλητές, οδηγούν σε χαμηλότερη αξιοπιστία των καταναλωτών (Berg et al., 2020).

Η εμπιστοσύνη των καταναλωτών στις διαδικτυακές αξιολογήσεις και κριτικές μπορεί να έχει πληγεί με την επικράτηση ψεύτικων και παραπλανητικών αξιολογήσεων και κριτικών. Πράγματι, πολλοί καταναλωτές έχουν δει αξιολογήσεις ή κριτικές που θεώρησαν ανέντιμες. Ομοίως, ορισμένοι καταναλωτές δυσκολεύονται να κάνουν διαφοροποίηση μεταξύ πωλητών/παρόχων επειδή υπάρχουν τόσοι πολλοί που έχουν υψηλές αξιολογήσεις (Poushneh et al., 2022).

2.4 Ζητήματα με διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών

Παρά τα πολλά οφέλη που μπορούν να αποφέρουν οι διαδικτυακές αξιολογήσεις και κριτικές καταναλωτών στους καταναλωτές, έχουν εντοπιστεί οι ακόλουθοι κίνδυνοι και προκλήσεις για τους καταναλωτές (Wang et al., 2020):

- Ψεύτικες αξιολογήσεις και κριτικές ή/και παραπλανητικές πρακτικές.
- Ανακριβείς αξιολογήσεις και κριτικές που δεν παρέχουν πάντα αντικειμενικά στοιχεία ποιότητας.
- Προκαταλήψεις των καταναλωτών κατά την παραγωγή και χρήση αξιολογήσεων.

Η αυξανόμενη χρήση διαδικτυακών αξιολογήσεων και κριτικών καταναλωτών για σκοπούς μάρκετινγκ και πωλήσεων έχει συνδυαστεί με ψευδείς και παραπλανητικές πρακτικές αναθεώρησης για την προώθηση αγαθών ή υπηρεσιών ή για την απαξίωση των ανταγωνιστών. Τέτοιες παραπλανητικές πρακτικές περιλαμβάνουν, για παράδειγμα (Wang et al., 2020):

- Αποτυχία γνωστοποίησης κριτικών με κίνητρα.
- Καταστολή αρνητικών σχολίων.
- Χειρισμός γνήσιων αξιολογήσεων καταναλωτών.

Τόσο οι ψεύτικες αξιολογήσεις και κριτικές όσο και οι παραπλανητικές πρακτικές παραμορφώνουν τις πληροφορίες για τους καταναλωτές και παραβιάζουν τη Σύσταση του ΟΟΣΑ για το ηλεκτρονικό εμπόριο.

2.5 Ψεύτικες βαθμολογίες και κριτικές

Ορισμένες χώρες έχουν εντοπίσει μια αυξανόμενη πρακτική ψεύτικων κριτικών που έχουν σχεδιαστεί για να μοιάζουν με ανεξάρτητες και αυθεντικές κριτικές καταναλωτών. Η παραγωγή μεγάλου αριθμού ευνοϊκών αξιολογήσεων και κριτικών που εμφανίζονται ως αυθεντικές εμπειρίες αμερόληπτων καταναλωτών («astroturfing») συχνά στοχεύει στην ενίσχυση των αξιολογήσεών τους ή στη μείωση των αξιολογήσεων των ανταγωνιστών τους. Ορισμένες επιχειρήσεις, για παράδειγμα, έχουν ενθαρρύνει τους δικούς τους υπαλλήλους να δημοσιεύουν θετικές αξιολογήσεις και κριτικές ή να δίνουν κίνητρα στους καταναλωτές να το κάνουν. Έρευνα του 2018 από τη βρετανική οργάνωση καταναλωτών What? δείχνει πώς οι καταναλωτές στρατολογούνται και ανταμείβονται για να γράφουν ψεύτικες κριτικές μέσω ομάδων μέσω κοινωνικής δικτύωσης. Ορισμένες επιχειρήσεις έχουν επίσης προσλάβει εταιρείες «βελτιστοποίησης φήμης» ή εταιρείες «βελτιστοποίησης μηχανών αναζήτησης» (search engine optimisation - SEO), οι οποίες πληρώνουν τρίτους για να δημοσιεύουν ψεύτικες διαδικτυακές αξιολογήσεις και κριτικές. Αυτές οι αξιολογήσεις και οι κριτικές είναι συχνά πλήρως κατασκευασμένες και κανένας πραγματικός καταναλωτής δεν έχει εμπλακεί με κανέναν τρόπο στην παραγωγή τους (Wang et al, 2020).

Ενώ οι ψεύτικες αξιολογήσεις και κριτικές χρησιμοποιούνται συχνά σε σχέση με παραπλανητικές ή δόλιες πρακτικές, όπως παγίδες συνδρομής και απάτες, μπορούν επίσης να αναπτυχθούν από νόμιμες επιχειρήσεις για να αποφέρουν οικονομικά οφέλη. Η παραγωγή ψεύτικων κριτικών μπορεί να είναι μια εκτεταμένη, καλά οργανωμένη και κερδοφόρα επιχείρηση. Ένας αριθμός μελετών έχει δημιουργήσει ευρέως ποικίλες εκτιμήσεις για τον όγκο των ψεύτικων κριτικών στην αγορά. Μια γαλλική μελέτη του 2015 εντόπισε μεταξύ 10% και 30% των κριτικών ως ψεύτικες, με έως και 45% από αυτές να είναι ύποπτες. Η EC θεωρεί τις ψεύτικες κριτικές ως έναν από τους παράγοντες που στρεβλώνουν περισσότερο την αγορά στον τομέα του ηλεκτρονικού εμπορίου (Plotkina et al., 2020).

Μελέτες δείχνουν ότι μπορούν να βρεθούν εύκολα στο διαδίκτυο οι επιχειρήσεις που παρέχουν υπηρεσίες ψεύτικων κριτικών και ότι ορισμένες ψεύτικες κριτικές

εξακολουθούν να περνούν τους μηχανισμούς επαλήθευσης των πλατφορμών χωρίς να εντοπιστούν. Παρόλο που υπάρχει λίγη διαθέσιμη έρευνα, μπορεί να υποθεθεί ότι οι κίνδυνοι των ψεύτικων κριτικών είναι υψηλότεροι στα «ανοικτά» συστήματα αναθεώρησης και ότι τα «κλειστά» συστήματα είναι πιο πιθανό να διαθέτουν διασφαλίσεις για την αποτροπή δημοσίευσης ψεύτικων κριτικών. Ωστόσο, εάν οι επιχειρήσεις των οποίων τα προϊόντα και οι υπηρεσίες υπόκεινται σε έλεγχο συμμετέχουν στην παραγωγή ψεύτικων κριτικών από μόνες τους, ψεύτικες κριτικές μπορούν επίσης να προκύψουν σε «κλειστά» συστήματα αξιολόγησης. Σε αυτήν την περίπτωση, θα περίμενε κανείς ότι η ζημιά των καταναλωτών θα ήταν πιο εκτεταμένη, καθώς οι καταναλωτές θα εμπιστεύονται γενικά περισσότερο τις αξιολογήσεις ενός «κλειστού» συστήματος. Αρκετές μελέτες δείχνουν ότι οι καταναλωτές δυσκολεύονται να αναγνωρίσουν ψεύτικες αξιολογήσεις και κριτικές (Ho & Hancock, 2019).

Μια άλλη μελέτη δείχνει ότι οι καταναλωτές στην πραγματικότητα υπερεκτιμούν την ικανότητά τους να αναγνωρίζουν τις ψεύτικες κριτικές. Οι καταναλωτές βασίζονται επίσης σε λογισμικό για τον εντοπισμό ψεύτικων κριτικών. Οι επιχειρήσεις που χρησιμοποιούν αξιολογήσεις και κριτικές, συμπεριλαμβανομένων των διαδικτυακών πλατφορμών, καταβάλλουν συνεχώς προσπάθειες για να ελαχιστοποιήσουν τον αριθμό των ψεύτικων αξιολογήσεων και κριτικών στους ιστότοπούς τους, μεταξύ άλλων μέσω της χρήσης αλγορίθμων και μηχανικής μάθησης, καθώς και μηχανισμών χειροκίνητου ελέγχου για τον εντοπισμό και τον αποκλεισμό ψεύτικων αξιολογήσεων και κριτικές. Ωστόσο, με τις προσπάθειες παράκαμψης των προστατευτικών τους μέτρων να εξελίσσονται συνεχώς, ορισμένες διαδικτυακές πλατφόρμες παραδέχονται ότι δεν μπορούν να εγγυηθούν ότι οι ιστοσελίδες τους δεν περιέχουν ψεύτικες κριτικές (Plotkina et al., 2020).

Η αντιμετώπιση του ζητήματος των ψεύτικων αξιολογήσεων και κριτικών απαιτεί νομοθεσία που να απαγορεύει τις παραπλανητικές και δόλιες πρακτικές. Η αυτορρύθμιση θα μπορούσε επίσης να συμβάλει σημαντικά στην εφαρμογή και την τήρηση των προτύπων από την αγορά. Επιπλέον, ενδέχεται να υπάρχουν οφέλη από την ανταλλαγή πληροφοριών και τη στενότερη συνεργασία μεταξύ των αρχών των καταναλωτών και των διαδικτυακών πλατφορμών όπου ενδείκνυται, όπως και ανάπτυξη συστημάτων πιστοποίησης τρίτων για αξιολογήσεις. Ωστόσο, η

πιστοποίηση, ως γενική απαίτηση για τη λειτουργία των συστημάτων αναθεώρησης, θα μπορούσε να δημιουργήσει εμπόδια εισόδου που ωφελούν τις υπάρχουσες επιχειρήσεις έναντι εκείνων που θέλουν να εισέλθουν στην αγορά (Plotkina et al., 2020).

Μέσω της ενημέρωσης, οι καταναλωτές θα μπορούσαν να ενημερωθούν σχετικά με τους κινδύνους των ψεύτικων αξιολογήσεων και κριτικών και να ενημερωθούν για το πώς να αναγνωρίζουν τις ψεύτικες αξιολογήσεις και κριτικές και να ενεργούν βάσει αυτών. Αυτό θα μπορούσε να αποτελέσει καθήκον για τις επιχειρήσεις, την κοινωνία των πολιτών καθώς και τις κυβερνήσεις (Plotkina et al., 2020).

2.6 Άλλες παραπλανητικές πρακτικές

Ορισμένες επιχειρήσεις έχουν χρησιμοποιήσει την αρχιτεκτονική και τις διοικητικές διαδικασίες των αξιολογήσεων και των κριτικών για να παραπλανήσουν τους καταναλωτές. Άλλες απέκλεισαν τις γνήσιες αρνητικές αξιολογήσεις και αξιολογήσεις, για παράδειγμα, διαγράφοντας, αποκλείοντας ή τιμωρώντας τις αρνητικές κριτικές. Άλλοι έχουν χειραγωγήσει το περιεχόμενο ή επηρέασαν τους καταναλωτές να το αλλάξουν. Τέτοιες πρακτικές με τη σειρά τους έχουν επηρεάσει τις αξιολογήσεις και τις κριτικές άλλων καταναλωτών. Οι επιχειρήσεις αντιμετώπισαν επίσης τις αρνητικές κριτικές διαφορετικά από τις θετικές, για παράδειγμα, αναβάλλοντας αδικαιολόγητα τη δημοσίευση αρνητικών κριτικών με το πρόσχημα της επανεξέτασής τους ή επίλυσης του προβλήματος για τον καταναλωτή (Fang et al., 2019).

Οι αρνητικές κριτικές έχουν επίσης αντιμετωπιστεί ως παράπονα ή ψευδείς ισχυρισμοί. Έχουν επίσης παρατηρηθεί ορισμένα παραδείγματα στα οποία οι επιχειρήσεις προσπάθησαν να δεσμεύσουν συμβατικά τους καταναλωτές να μην αφήνουν αρνητικά σχόλια. Ειδικότερα, οι διευκολυντές αξιολόγησης τρίτων (επιχειρήσεις τύπου 3) θα μπορούσαν να έχουν πρόσθετα κίνητρα για να αλλοιώσουν το περιεχόμενο της κριτικής καταναλωτών. Τα επιχειρηματικά τους μοντέλα εξαρτώνται από την επιτυχία τους να παράγουν θετικά σχόλια για τους πελάτες τους για να αποκτήσουν μελλοντικές επιχειρήσεις. Η διασφάλιση της ακεραιότητας των αξιολογήσεων και των κριτικών των

καταναλωτών μπορεί να μην είναι πάντα κορυφαία προτεραιότητα για αυτές τις επιχειρήσεις (Fang et al., 2019).

2.6.1 Κριτικές με κίνητρα

Η προσφορά ανταμοιβών στους καταναλωτές για τις αξιολογήσεις ή τις κριτικές τους είναι ένας τρόπος για τις επιχειρήσεις να αυξήσουν τον αριθμό περισσότερων (θετικών) αξιολογήσεων και κριτικών. Οι καταναλωτές μπορούν να ανταμειφθούν με διαφορετικούς τρόπους, μεταξύ άλλων μέσω, για παράδειγμα, χρηματικής πληρωμής, εκπτώσεων, πιστώσεων και άλλων οικονομικών οφελών. Η παροχή «δωρεάν» αγαθών ή υπηρεσιών, μπορεί επίσης να χρησιμοποιηθεί για να παρακινήσει τους καταναλωτές να δημοσιεύουν κριτικές. Οι κριτικές με κίνητρα σε ιστολόγια, διαδικτυακά βίντεο και πλατφόρμες μέσω κοινωνικής δικτύωσης παρατηρούνται όλο και περισσότερο και αποτελούν βασικό στοιχείο για τη διαφήμιση προϊόντων. Μπορούν να εγείρουν ζητήματα για τους καταναλωτές που μπορεί να δυσκολεύονται να προσδιορίσουν εάν οι αναρτήσεις που γίνονται σε μέσα κοινωνικής δικτύωσης, ιστολόγια ή σε άλλα περιβάλλοντα αντικατοπτρίζουν αμερόληπτες απόψεις ή πληρωμένη διαφήμιση. Αυτό έχει γίνει ιδιαίτερα προβληματικό δεδομένου ότι πολλές επιχειρήσεις συχνά παρέχουν δωρεάν προϊόντα σε άτομα που επηρεάζουν τα μέσα κοινωνικής δικτύωσης (Wang et al., 2020).

Έχουν εκφραστεί ανησυχίες σχετικά με το γεγονός ότι πολλοί διαφημιστές και πάροχοι διαδικτυακού περιεχομένου, όπως οι influencers μέσω κοινωνικής δικτύωσης, δεν αποκαλύπτουν πληρωμές ή άλλες εμπορικές ρυθμίσεις ξεκάθαρα και εμφανώς, καθιστώντας δύσκολο για τους καταναλωτές να εκτιμήσουν το βάρος και την αξιοπιστία των προϊόντων. Μια μελέτη έδειξε ότι οι κριτικές γίνονται πιο ευνοϊκές για την επιχείρηση που προσφέρει το κίνητρο. Οι αγοραστές ανταποδίδουν τις «καλές πράξεις» των πωλητών με συχνότερη και πιο ευνοϊκή ανατροφοδότηση. Με άλλα λόγια: οι πωλητές μπορούν να αγοράσουν ανατροφοδότηση, αλλά αυτή η ανατροφοδότηση είναι πιθανό να είναι προκατειλημμένη. Αυτό μπορεί να οδηγήσει σε μείωση της αυθεντικότητας και της αμεροληψίας των διαδικτυακών κριτικών καταναλωτών, γεγονός που μπορεί να επηρεάσει την εμπιστοσύνη των καταναλωτών (Wang et al., 2020).

Για να αποφευχθούν τα κίνητρα από μεροληπτικές αξιολογήσεις και κριτικές, προτείνεται οι αξιολογήσεις και οι κριτικές να επικεντρώνονται σε σχετικές πληροφορίες για το προϊόν ή τον πωλητή και όχι σε συναισθήματα. Σύμφωνα με αυτήν την άποψη, οι καταναλωτές θα γνωρίζουν ότι μπορούν να εμπιστευτούν τις επιχειρήσεις που δίνουν κίνητρα σε κριτικές για να αποκτήσουν μόνο έγκυρες, σχετικές πληροφορίες προϊόντος. Δεν είναι σαφές, ωστόσο, εάν οι καταναλωτές θα έβρισκαν χρήσιμες τέτοιες κριτικές, καθώς η αξία των κριτικών καταναλωτών είναι ότι περιέχουν υποκειμενικές εκτιμήσεις. Ένα θετικό αποτέλεσμα της παροχής κινήτρων για ανατροφοδότηση φήμης θα μπορούσε να είναι ότι οι καταναλωτές με πιο μετριοπαθείς απόψεις θα παρέχουν ανατροφοδότηση, γεγονός που θα μπορούσε να αυξήσει τον αριθμό των καταναλωτών που αξιολογούν, μειώνοντας την πόλωση των αξιολογήσεων και των κριτικών (Marinescu et al., 2018).

Ένα πιθανό μειονέκτημα θα μπορούσε να είναι ότι η καθιερωμένη θετική φήμη μπορεί να λειτουργήσει ως εμπόδιο εισόδου στην αγορά για νέες επιχειρήσεις που δεν έχουν ακόμη διαδικτυακή φήμη. Ωστόσο, οι αγορές για αξιολογήσεις θα μπορούσαν ενδεχομένως να συμβάλουν στην άμβλυνση ορισμένων από αυτά τα εμπόδια. Στην πράξη, πολλές μεγάλες διαδικτυακές πλατφόρμες δεν δέχονται αξιολογήσεις που βασίζονται σε κίνητρα που δεν αποκαλύπτονται (Wang et al., 2020).

2.6.2 Συγκεντρωτικές αξιολογήσεις και άλλες αναδυόμενες πρακτικές

Πολλοί ιστότοποι και διαδικτυακές πλατφόρμες παράγουν συγκεντρωτικές αξιολογήσεις για προϊόντα ή πωλητές. Αυτές οι αξιολογήσεις συνήθως αποτελούνται από τον αριθμό των θετικών κριτικών ως ποσοστό του συνολικού αριθμού των κριτικών. Λόγω του τρόπου που δημιουργούνται οι βαθμολογίες, είναι γενικά υψηλές και παρουσιάζουν μικρή κατανομή - μια επίδραση που ονομάζεται βαθμολογικός πληθωρισμός. Οι αξιολογήσεις τείνουν να διογκώνονται σε όλες τις διαδικτυακές αγορές. Ορισμένοι μελετητές αμφισβητούν το εάν οι καταναλωτές είναι σε θέση να κάνουν χρήσιμες διακρίσεις μεταξύ καλών και λιγότερο καλών πωλητών ή προϊόντων, με βάση τις διογκωμένες αξιολογήσεις (Schoenmüller et al., 2018).

Οι καταναλωτές μπορεί επίσης να παραπλανηθούν όταν οι συγκεντρωτικές βαθμολογίες βασίζονται σε αξιολογήσεις με κίνητρα ή ίδιες κριτικές που δημοσιεύονται σε διαφορετικές πλατφόρμες. Ένας πιθανός τρόπος αντιμετώπισης του

πληθωρισμού θα μπορούσε να είναι η προσθήκη του αριθμού των συναλλαγών που δεν εξετάστηκαν από τους αγοραστές (οι λεγόμενες «σιωπηλές κριτικές») στην εξίσωση που καθορίζει τη συγκεντρωτική βαθμολογία του πωλητή. Κατά τη διάρκεια ενός πειράματος, αυτές οι εναλλακτικές αξιολογήσεις ήταν σε θέση να προβλέψουν εάν ένας αγοραστής επέστρεψε για να αγοράσει ξανά από τον ίδιο πωλητή (Kusuma et al., 2020).

Ορισμένες διαδικτυακές πλατφόρμες όχι μόνο επιτρέπουν στον καταναλωτή να παρέχει σχόλια για τον πωλητή, αλλά και στον πωλητή να βαθμολογεί τον καταναλωτή (π.χ. Airbnb και Uber). Ωστόσο, ανάλογα με την αρχιτεκτονική του συστήματος, ένας μηχανισμός αμφίπλευρης ανάδρασης θα μπορούσε να παράγει αντίποινα που είναι πιθανό να διογκώσουν τις αξιολογήσεις των καταναλωτών και την αναθεώρηση των προϊόντων ή/και των υπηρεσιών του πωλητή. Η αρνητική βαθμολογία ενός πωλητή μπορεί να επηρεάσει τη μελλοντική προθυμία του καταναλωτή να συναλλάσσεται με τον πωλητή, την ποιότητα και την τιμή των προϊόντων ή/και των υπηρεσιών. Στην πραγματικότητα, τα αμφίπλευρα συστήματα ανατροφοδότησης θα μπορούσαν να τιμωρήσουν τις αρνητικές αξιολογήσεις και κριτικές (Kusuma et al., 2020).

Μια λύση σε αυτό το ζήτημα θα μπορούσε να είναι ο καταναλωτής και ο πωλητής να αξιολογήσουν ο ένας τον άλλο χωρίς να γνωρίζουν το περιεχόμενο της αξιολόγησης ή/και της κριτικής του άλλου. Ένα σχετικό αναδυόμενο ζήτημα που επηρεάζει τη χρησιμότητα των κριτικών και των αξιολογήσεων για τους καταναλωτές προκύπτει στο πλαίσιο της παραδοσιακής διαφήμισης. Σήμερα, είναι ευρέως αποδεκτό ότι οι επιχειρήσεις χρησιμοποιούν αξιολογήσεις και κριτικές ως μορφή διαφήμισης των προϊόντων και των υπηρεσιών τους. Όλο και περισσότερο, οι διαφημιζόμενοι διαφημίζουν κριτικές καταναλωτών στη διαφήμιση, προβάλλοντας ισχυρισμούς όπως "Πιο συνιστώμενες" ή "Περισσότερες αξιολογήσεις με 5 αστέρια". Στις Ηνωμένες Πολιτείες, ένας αυτορυθμιζόμενος οργανισμός, το National Advertising Division of the Better Business Bureau (NAD) έχει αντιμετωπίσει αυτήν την πρακτική, παρέχοντας καθοδήγηση σχετικά με την εφαρμογή των παραδοσιακών αρχών της αλήθειας στη διαφήμιση σε αυτό το πλαίσιο. Σύμφωνα με το NAD, τα δεδομένα αναθεώρησης πρέπει να περιορίζονται σε κριτικές που σχετίζονται με επαληθευμένες αγορές και αντιπροσωπεύουν τις πραγματικές απόψεις των καταναλωτών. Επιπλέον, οι

αξιολογήσεις που συγκεντρώνονται σε πολλές πλατφόρμες θα πρέπει να ελέγχονται για να διασφαλιστεί ότι η ίδια κριτική δεν μετράται πολλές φορές (Kaur & Singla, 2019).

2.7 Ο ρόλος που διαδραματίζουν οι διαδικτυακές πλατφόρμες

Σε σχέση με τις παραπλανητικές πρακτικές αναθεώρησης, υπάρχουν ζητήματα πολιτικής σχετικά με το ρόλο και τις ευθύνες των διαδικτυακών πλατφορμών στη διασφάλιση της αυθεντικότητας και της αμεροληψίας των διαδικτυακών αξιολογήσεων και κριτικών καταναλωτών. Σε υψηλό επίπεδο, υπάρχει συναίνεση ότι οι διαδικτυακές πλατφόρμες μπορούν να διαδραματίσουν ρόλο στην προώθηση των αρχών της αλήθειας στη διαφήμιση. Αυτό θα μπορούσε να περιλαμβάνει, για παράδειγμα, την εφαρμογή και την επιβολή εσωτερικών κανόνων στους παρόχους περιεχομένου σε διαδικτυακές πλατφόρμες, πχ να αποκαλύπτουν ξεκάθαρα και εμφανώς υλικές συνδέσεις με τους διαφημιστές στις αξιολογήσεις τους (Plotkina et al., 2020).

Οι διαδικτυακές πλατφόρμες θα μπορούσαν επίσης να διαδραματίσουν ρόλο στην ανάπτυξη καλύτερου σχεδιασμού αξιολογήσεων μέσω, για παράδειγμα, αυξημένης και δίκαιης χρήσης μηχανισμών αμφίπλευρης ανάδρασης. Επιπλέον, για να μπορέσουν οι καταναλωτές να εντοπίσουν καλούς πωλητές ή προϊόντα, οι διαδικτυακές πλατφόρμες θα μπορούσαν να αναλύσουν και να παρέχουν άλλες ποιοτικές πληροφορίες για πωλητές ή προϊόντα, όπως πληροφορίες από καταγγελίες καταναλωτών. Το λογισμικό μηχανικής εκμάθησης μπορεί επίσης να βοηθήσει στη συλλογή των σχετικών πληροφοριών για το σκοπό αυτό (Fang et al., 2019).

2.8 Η ακρίβεια των αξιολογήσεων και των κριτικών των καταναλωτών ως ενδείξεις ποιότητας

Όπως περιγράφεται στην προηγούμενη ενότητα, η ακρίβεια των αξιολογήσεων και των κριτικών των καταναλωτών στο διαδίκτυο μπορεί να «πάσχει» από σκόπιμη εξαπάτηση ή χειραγώγηση. Αυτή η ενότητα εξετάζει ορισμένους από τους τρόπους με τους

οποίους η ανακρίβεια των αξιολογήσεων μπορεί να προκύψει από μη εσκεμμένες ενέργειες ή ελαττώματα στον τρόπο λειτουργίας των συστημάτων αξιολόγησης. Αυτό το θέμα έχει γενικά προσελκύσει λιγότερη προσοχή σε επίπεδο πολιτικής από το ζήτημα των ψεύτικων ή παραπλανητικών αξιολογήσεων και κριτικών. Ακόμα κι αν οι διαδικτυακές αξιολογήσεις και κριτικές δεν είναι απολύτως ακριβείς, μπορούν να βοηθήσουν τους καταναλωτές να συγκρίνουν τις σχετικές αξίες των πωλητών ή των αγαθών και των υπηρεσιών. Ωστόσο, θα ήταν ωφέλιμο για τις αξιολογήσεις και τις κριτικές να έχουν ένα σημαντικό επίπεδο ακρίβειας σχετικά με τη χρηστικότητα και την ποιότητα των προϊόντων ή των υπηρεσιών. Οι περισσότερες μελέτες συμφωνούν ότι η συντριπτική πλειονότητα των κριτικών των καταναλωτών γράφεται από περίπου το 1% όλων των καταναλωτών. Είναι πιθανό ότι αυτή η ομάδα καταναλωτών δεν αντιπροσωπεύει μια μέση ομάδα καταναλωτών και μπορεί να μην αντικατοπτρίζει ή να αντιμετωπίζει πάντα τις σχετικές πτυχές ποιότητας για το υπόλοιπο 99% των καταναλωτών (Fang et al., 2019).

Δεύτερον, τα μεγέθη δειγμάτων διατρέχουν τον κίνδυνο να είναι πολύ χαμηλά για να παρέχουν ακριβή σήματα ποιότητας και οι καταναλωτές μπορεί να μην το γνωρίζουν απαραίτητα (De Langhe et al., 2016). Ένας πιθανός τρόπος αύξησης του αριθμού των καταναλωτών που αξιολογούν θα μπορούσε να είναι η αποστολή υπενθυμίσεων ή η παροχή ανταμοιβών (Schoenmüller et al., 2018)

Ένας άλλος λόγος που οι κριτικές των καταναλωτών μπορεί να μην αντιπροσωπεύουν την εμπειρία του μέσου καταναλωτή ενός αγαθού ή μιας υπηρεσίας είναι ότι οι καταναλωτές είναι πιο πιθανό να γράφουν κριτικές είτε με πολύ θετικούς είτε με αρνητικούς τρόπους, τη λεγόμενη «πόλωση» των κριτικών. Επιπλέον, η έρευνα δείχνει ότι οι κίνδυνοι ανακριβών κριτικών καταναλωτών μπορεί να είναι υψηλότεροι σε αγορές με εμπειρικά αγαθά και υπηρεσίες που αγοράζονται με βάση υποκειμενικές πτυχές ποιότητας. Οι καταναλωτές συχνά παράγουν κριτικές αμέσως μετά την αγορά ενός αγαθού ή μιας υπηρεσίας. Επομένως, ειδικά για αγαθά και υπηρεσίες με μεγαλύτερη διάρκεια ζωής, οι κριτικές ενδέχεται να μην παρέχουν αξιόπιστη ένδειξη ποιότητας για μακροπρόθεσμη βάση (π.χ. ανθεκτικότητα). Ζητώντας από τους καταναλωτές μια ενημέρωση μιας αναθεώρησης μετά από ένα ορισμένο χρονικό διάστημα θα μπορούσε να εγγυηθεί πιο πολύτιμες πληροφορίες. Υπάρχει επίσης ο

κίνδυνος οι καταναλωτές, ιδιαίτερα όσοι ασχολούνται με ηλεκτρονικό εμπόριο μέσω διαδικτυακών πλατφορμών, να μην είναι πάντα σε θέση να διακρίνουν τις κριτικές σχετικά με μια συναλλαγή (π.χ. χρόνοι παράδοσης και χρεώσεις) και τις κριτικές σχετικά με τη χρήση ή την ποιότητα ενός προϊόντος (Schoenmüller et al., 2018).

2.9 Προκαταλήψεις συμπεριφοράς των καταναλωτών στην παραγωγή και χρήση αξιολογήσεων και κριτικών

Κατά την παραγωγή αξιολογήσεων και κριτικών, οι καταναλωτές φαίνεται να επηρεάζονται από διάφορες συμπεριφορικές προκαταλήψεις. Μια τέτοια προκατάληψη είναι η κοινωνική επιρροή. Σύμφωνα με μια μελέτη του 2014, οι καταναλωτές τείνουν να αξιολογούν τα αγαθά και τις υπηρεσίες πιο θετικά εάν υπάρχουν ήδη διαθέσιμες ορισμένες θετικές αξιολογήσεις και κριτικές. Αυτό φαίνεται ιδιαίτερα αληθές εάν αυτές οι αξιολογήσεις και κριτικές προέρχονται από φίλους ή συγγενείς. Ορισμένες μελέτες διαπίστωσαν επίσης ότι οι καταναλωτές τείνουν να παρέχουν ανατροφοδότηση με ακραία κατανομή και να στρέφονται προς τα θετικά, με την αριθμητική ανατροφοδότηση (βαθμολογήσεις) να επηρεάζεται περισσότερο από την αφηγηματική ανατροφοδότηση (κριτικές). Οι αξιολογήσεις και οι κριτικές χάνουν την πληροφοριακή τους ικανότητα ως αποτέλεσμα της ακραίας κατανομής (Schoenmüller et al., 2018).

Μια άλλη μελέτη αναφέρει ότι οι καταναλωτές τείνουν να δημοσιεύουν πιο πολωμένες αξιολογήσεις και κριτικές, καθώς θέλουν να επηρεάσουν τις αξιολογήσεις και τις κριτικές ενός προϊόντος προκειμένου να τις ευθυγραμμίσουν με τη δική τους γνώμη (προκατάληψη αυτοεπιλογής). Μια μελέτη του 2018, η οποία επιβεβαιώνει τη μεροληψία επιλογής, προτείνει ότι τα (νομισματικά) κίνητρα θα μπορούσαν να μειώσουν τις επιπτώσεις της μεροληψίας επιλογής προσελκύοντας αξιολογήσεις και κριτικές από καταναλωτές με μέτριες απόψεις (Marinescu et al., 2018).

Η διαδικτυακή αλληλεπίδραση των καταναλωτών με τους πωλητές επηρεάζει την αξιολόγησή τους για ένα προϊόν. Οι καταναλωτές που συζητούν τα συγκεκριμένα χαρακτηριστικά ενός προϊόντος τείνουν να το εκτιμούν περισσότερο. Η ίδια μελέτη

διαπίστωσε επίσης ότι οι αξιολογήσεις και οι κριτικές των καταναλωτών αντικατοπτρίζουν εν μέρει τις διαφορές μεταξύ της πραγματικής ποιότητας του προϊόντος και των προηγούμενων προσδοκιών, όπως συνάγεται από προηγούμενες αξιολογήσεις. Η τάση των καταναλωτών να εκτιμούν τις ακραίες κριτικές είναι ο κύριος μοχλός για την ευρεία κατανομή των αξιολογήσεων. Επιπλέον, μια μελέτη της EC του 2017 για αγορές ομότιμων πλατφορμών υποδηλώνει ότι μόνο το 20% των χρηστών άφησαν αρνητική κριτική ή βαθμολογία αφού αντιμετώπισαν ένα πρόβλημα συναλλαγών, υπονοώντας πιθανές προκαταλήψεις στην παραγωγή αξιολογήσεων και κριτικών καταναλωτών στο διαδίκτυο (Schoenmüller et al., 2018).

Παρόμοιες προκαταλήψεις συμπεριφοράς έχουν παρατηρηθεί σε σχέση με τη χρήση αξιολογήσεων και κριτικών από τους καταναλωτές. Από τη μία πλευρά, οι καταναλωτές φαίνεται να επηρεάζονται περισσότερο από ακραίες αξιολογήσεις και κριτικές, τόσο θετικές όσο και αρνητικές. Από την άλλη πλευρά, ορισμένες έρευνες δείχνουν ότι οι καταναλωτές είναι πιο πιθανό να απορρίψουν τις ακραίες κριτικές ως ανέντιμες ή ψεύτικες. Οι καταναλωτές τείνουν επίσης να εκτιμούν τις αρνητικές κριτικές περισσότερο από τις θετικές κριτικές. Τέλος, η έρευνα προτείνει ότι οι αφηγηματικές κριτικές φαίνεται να έχουν μεγαλύτερο αντίκτυπο από τις μη αφηγηματικές (π.χ. στατιστικές) κριτικές. Μια μελέτη με στόχο τον εντοπισμό των παραγόντων που επηρεάζουν την αποτελεσματικότητα των διαδικτυακών προφορικών μηνυμάτων - συμπεριλαμβανομένων των διαδικτυακών κριτικών και αξιολογήσεων των καταναλωτών - παρατήρησε ότι η έρευνα για το θέμα παραμένει σε μεγάλο βαθμό κατακερματισμένη και ότι τα ευρήματα μερικές φορές είναι ασυνεπή και ακόμη και αντιφατικά. Ωστόσο, υπάρχει σαφής ένδειξη ότι οι καταναλωτές είναι περισσότερο ή λιγότερο προκατειλημμένοι όταν παρέχουν και χρησιμοποιούν αξιολογήσεις και κριτικές, αν και είναι λιγότερο σαφές με ποιους ακριβώς τρόπους και σε ποιο βαθμό (Qahri-Saremi & Montazemi, 2019).

Συνολικά, οι αξιολογήσεις και οι κριτικές των καταναλωτών στο διαδίκτυο αποφέρουν μια σειρά από οφέλη τόσο στους καταναλωτές όσο και στις επιχειρήσεις. Για τους καταναλωτές, μπορούν να οδηγήσουν σε καλύτερα ενημερωμένες αποφάσεις αγοράς, μειωμένο κόστος αναζήτησης, χαμηλότερες τιμές και αύξηση της ποιότητας των προϊόντων και των υπηρεσιών. Διευκολύνουν τη συμμετοχή των καταναλωτών στο

ηλεκτρονικό εμπόριο και συμβάλλουν στην οικοδόμηση εμπιστοσύνης μεταξύ των καταναλωτών και των επιχειρήσεων. Οι καταναλωτές γενικά βασίζονται στις διαδικτυακές αξιολογήσεις και κριτικές των καταναλωτών στις αποφάσεις αγοράς τους, τόσο στο διαδίκτυο όσο και διά ζώσης. Οι επιχειρήσεις, από την πλευρά τους, χρησιμοποιούν τις αξιολογήσεις και τις κριτικές των καταναλωτών ως εργαλείο για να βελτιώνουν τα προϊόντα τους σε συνεχή βάση και να βοηθούν στη διαχείριση της φήμης τους. Ως αποτέλεσμα, ορισμένες επιχειρήσεις έχουν αναπτύξει μηχανισμούς για τη λήψη ευνοϊκότερων αξιολογήσεων από τους καταναλωτές. Αυτό οδήγησε στην ανάπτυξη μιας νέας αγοράς στην οποία οι επιχειρήσεις παρέχουν σε άλλες επιχειρήσεις υπηρεσίες αξιολόγησης (Wang et al., 2020).

Από την άποψη της προστασίας των καταναλωτών, ορισμένες πτυχές σχετικά με τη λειτουργία των διαδικτυακών αξιολογήσεων και κριτικών καταναλωτών είναι προβληματικές και θα μπορούσαν, σε ορισμένες περιπτώσεις, να προκαλέσουν βλάβη στους καταναλωτές (Wang et al., 2020):

-Ψεύτικες και παραπλανητικές πρακτικές αναθεώρησης, οι οποίες επηρεάζουν αρνητικά τόσο τη λήψη αποφάσεων όσο και την εμπιστοσύνη των καταναλωτών.

-Ανακριβείς κριτικές καταναλωτών, οι οποίες δεν παρέχουν πάντα αντικειμενικά στοιχεία ποιότητας.

-Προκαταλήψεις συμπεριφοράς των καταναλωτών κατά την παραγωγή και χρήση διαδικτυακών αξιολογήσεων και κριτικών.

Για την αντιμετώπιση αυτών των ζητημάτων, διάφοροι οργανισμοί έχουν εφαρμόσει μέτρα για τον μετριασμό των (πιθανών) αρνητικών επιπτώσεων των διαδικτυακών αξιολογήσεων των καταναλωτών, όπως νομοθεσία, οδηγίες, μέτρα επιβολής, κοινοτικές οδηγίες και εκπαίδευση των καταναλωτών. Τέτοια μέτρα υποδεικνύουν ότι ως βασική απαίτηση, η διαχείριση των συστημάτων αξιολόγησης πρέπει να γίνεται με δίκαιο και διαφανή τρόπο. Θα πρέπει να παρέχουν πληροφορίες για τις γνώσεις εμπειρίες των καταναλωτών και οι επιχειρήσεις θα πρέπει να είναι διαφανείς σχετικά με τους τρόπους με τους οποίους δίνουν κίνητρα, επεξεργάζονται, φιλτράρουν και δημοσιεύουν κριτικές. Οι διαδικτυακές πλατφόρμες μπορούν να διαδραματίσουν μείζονος σημασίας ρόλο στη διατήρηση αυθεντικών και αμερόληπτων κριτικών. Οι

νόμοι περί καταναλωτών σε πολλές δικαιοδοσίες απαιτούν υποχρεωτικές γνωστοποιήσεις από τις επιχειρήσεις σχετικά με τις ουσιώδεις διασυνδέσεις μεταξύ μιας επιχείρησης και ενός αναθεωρητή (π.χ. γνωστοποίηση σχετικά με το γεγονός ότι μια κριτική υποκινήθηκε ή υποβλήθηκε σε επεξεργασία) (Wang et al., 2020).

Αν και δεν υπάρχουν πολλά στοιχεία σχετικά με τον αντίκτυπο ή την αποτελεσματικότητα αυτών των γνωστοποιήσεων, οι περισσότερες εταιρείες καταναλωτών ωστόσο ισχυρίζονται ότι τέτοιες γνωστοποιήσεις επηρεάζουν τη βαρύτητα και την αξιοπιστία που δίνει ένας καταναλωτής σε μια κριτική. Αν και υπάρχει ένα ευρύ φάσμα ερευνών σχετικά με τον αποτελεσματικό σχεδιασμό αποκάλυψης, θα ήταν χρήσιμη πρόσθετη έρευνα σχετικά με την αποτελεσματικότητα και τον βέλτιστο σχεδιασμό των γνωστοποιήσεων για διαδικτυακές αξιολογήσεις και κριτικές. Υπάρχουν επίσης λίγα εμπειρικά στοιχεία σχετικά με την αποτελεσματικότητα των διαφόρων προτύπων και απαιτήσεων που στοχεύουν στη βελτίωση των επιχειρηματικών πρακτικών που σχετίζονται με τις αξιολογήσεις και τις κριτικές των καταναλωτών στο διαδίκτυο. Περαιτέρω έρευνα σχετικά με τον βαθμό στον οποίο οι επιχειρήσεις συμμορφώνονται με αυτά τα πρότυπα και απαιτήσεις και βελτιώνουν πραγματικά την ποιότητα των διαδικτυακών αξιολογήσεων και κριτικών των καταναλωτών θα μπορούσε να είναι επωφελής (Wang et al., 2020).

ΚΕΦΑΛΑΙΟ 3^ο

3.1 Τι είναι το crowdfunding

Το Astroturfing αναφέρεται σε καμπάνιες που φαίνεται να διευθύνονται από συμμετέχοντες, αλλά στην πραγματικότητα υποστηρίζεται από σκόπιμα καλυμμένους χορηγούς. Το crowdfunding μπορεί να θεωρηθεί ως ένας συγκεκριμένος τύπος astroturfing, του οποίου οι συμμετέχοντες στην εκστρατεία οργανώνονται μέσω crowdsourcing. Η μελέτη του astroturfing μπορεί να ανιχνευθεί πίσω στο 1985, όταν οι ασφαλιστικές εταιρείες προσέλαβαν «astroturfs» για να ξεκινήσουν εκστρατείες που

προωθούν τα ενδιαφέροντά τους. Ξεκινώντας από το 1998, οι ερευνητές επικεντρώθηκαν στη μέτρηση της διαφοράς μεταξύ των εκστρατειών που ηγούνται από μία αληθινή βάση και astroturfs. Ένα άλλο παράδειγμα astroturfing είναι οι εκστρατείες που ξεκίνησαν από τη βιομηχανία πούρων στα τέλη της δεκαετίας του 1990. Με την αύξηση των φόρων στον καπνό και τους περισσότερους κανονισμούς για το κάπνισμα στις Ηνωμένες Πολιτείες, η βιομηχανία πούρων αντιμετώπισε δύσκολες στιγμές από τις αρχές της δεκαετίας του 1990. Προκειμένου να αποφευχθεί η απώλεια εισοδήματος, οι καπνοβιομηχανίες, μαζί με την Εθνική Συμμαχία Καπνιστών, ξεκίνησαν μια επιθετική εκστρατεία για την προστασία των συμφερόντων τους. Οι astroturfs έστειλαν κάρτες και επιστολές για να υπερασπιστούν τα δικαιώματα των καπνιστών (Givel, 2007).

Με την ανάπτυξη των τεχνολογιών πληροφοριών και επικοινωνιών, το astroturfing έχει εξελίξει τον τρόπο προσέγγισης των ανθρώπων. Εκτός από τις επιστολές και τις κάρτες, χρησιμοποιούνται αυτοματοποιημένες τηλεφωνικές κλήσεις, ιστότοποι και μηνύματα ηλεκτρονικού ταχυδρομείου για να γίνει το astroturfing παραγωγικό και οικονομικό. Το 2001, το Americans for Technology Leadership (ATL) και η Ένωση για την Ανταγωνιστική Τεχνολογία ξεκίνησαν τους υποστηρικτές τους να κάνουν τους ανθρώπους να στέλνουν επιστολές στους γενικούς εισαγγελείς της πολιτείας τους. Οι επιστολές ήταν εκ των προτέρων γραμμένες και είχαν σκοπό να πείσουν τους γενικούς εισαγγελείς να αποσύρουν μήνυση κατά της Microsoft. Αργότερα ανακαλύφθηκε ότι η ATL χρηματοδοτήθηκε σε μεγάλο βαθμό από τη Microsoft (Wu & Liu, 2018).

Το crowdfunding έγινε πιο διαδεδομένο γύρω στο 2010, επωφελούμενο από τη διαθεσιμότητα της οικονομικής οργάνωσης μεγάλου αριθμού συμμετεχόντων. Το χαμηλότερο κόστος του crowdfunding επιτρέπει επίσης σε μικρές ομάδες να ξεκινήσουν τις καμπάνιες τους. Σε πλατφόρμες του Διαδικτύου, όπως φόρουμ, blogs και μέσα κοινωνικής δικτύωσης, διαπιστώθηκε ότι το crowdfunding εμπλέκεται με πολιτικές και εμπορικές εκστρατείες. Προκειμένου να ανιχνευθεί το astroturfing, οι παραδοσιακές μέθοδοι βασίζονται στον χειροκίνητο έλεγχο του περιεχομένου. Για παράδειγμα, οι επιστολές από τις εκστρατείες ATL και Association for Competitive Technology βρέθηκαν να είναι παρόμοιες μεταξύ τους. Οι επιστολές γράφτηκαν ακόμη και από ανθρώπους που πέθαναν πριν υπογραφούν. Ωστόσο, με τη διαθεσιμότητα του

crowdsourcing, ο χειροκίνητος έλεγχος είναι πολύ δαπανηρός για να αντιμετωπίσει το μεγάλης κλίμακας crowdurfing (Menn & Sanders, 2008).

Σύμφωνα με μια μελέτη σε ιστότοπους crowdurfing, μια καμπάνια crowdurfing μπορεί να αποτελείται από 200 εργασίες. Επομένως, θα ήταν εξαιρετικά δύσκολο να βασιστούμε απλώς στον χειροκίνητο έλεγχο για να ανιχνεύσουμε τον τεράστιο όγκο πληροφοριών που προκαλεί το crowdurfing (Ratkiewicz et al., 2010).

Το crowdurfing, η σκοτεινή πλευρά και συνήθως απαρατήρητη πτυχή του crowdsourcing, αντιπροσωπεύει μια μορφή κυβερνοπαραπλάνησης στην οποία οι εργαζόμενοι πληρώνονται για να εκφράσουν μια ψευδή ψηφιακή γνώμη. Αν και μια τέτοια συμπεριφορά μπορεί να μην τιμωρείται βάσει της δικαιοδοσίας του επίσημου δικαίου, οι συνέπειές της είναι καταστροφικές για τη συνοχή και την αξιοπιστία των διαδικτυακών πληροφοριών. Η μελέτη των Rinta-Kahila & Soliman (2017) εξέτασε την τρέχουσα βιβλιογραφία για το θέμα και έθεσε τα θεμέλια για ένα θεωρητικό πλαίσιο που εξηγεί τη συμπεριφορά του crowdurfing. Οι ίδιοι ανέλυσαν το crowdurfing μέσα από τρεις ηθικές κανονιστικές προσεγγίσεις: την παραδοσιακή φιλοσοφική ηθική, την επιχειρηματική ηθική και τους κωδικοποιημένους κανόνες, κάνοντας την εφαρμογή σε ένα ενδεικτικό παράδειγμα διαδικτυακής πλατφόρμας που ενορχηστρώνει το εμπόριο κριτικών βιβλίων επί πληρωμή στο Amazon.

Αξίζει να επισημανθεί πως η μελέτη των Rinta-Kahila & Soliman (2017) συνέβαλε στη θεωρία εξηγώντας την ηθική λογική πίσω από το crowdurfing από την οπτική γωνία των βασικών παραγόντων που εμπλέκονται. Υποστηρίζεται ότι ενώ το crowdurfing δεν μπορεί να «αντέξει» μια κριτική εξέταση μέσω των δεοντολογικών ή κοινωνικών συμβολαίων, η τελολογική λογική, η θεωρία των μετόχων ή ορισμένα επίπεδα κωδικοποιημένων κανόνων μπορούν να επιτρέψουν στους φορείς που εμπλέκονται στην επιχείρηση να λειτουργούν με καθαρή συνείδηση. Η αυξημένη κατανόηση της συμπεριφοράς μπορεί να βοηθήσει τόσο τις πλατφόρμες των θυμάτων όσο και την κοινότητα του Διαδικτύου γενικότερα να καταπολεμήσουν αυτήν την πρόκληση.

Είναι αποδεδειγμένο ότι ο απώτερος στόχος της ασφάλειας των πληροφοριών είναι η προστασία τριών θεμελιωδών ποιότητων πληροφόρησης: ακεραιότητα, διαθεσιμότητα

και εμπιστευτικότητα (Loch et al., 1992) - κάτι που ο Mason (1986) περιέγραψε ως τέσσερα εξαιρετικά σημαντικά ζητήματα στην εποχή της πληροφορίας και τα ονόμασε PAPA (συντομογραφία για το privacy, accuracy, property, accessibility – δηλαδή, απόρρητο, ακρίβεια, ιδιοκτησία, προσβασιμότητα). Το crowdurfing είναι νέο φαινόμενο του Διαδικτύου που θέτει σε κίνδυνο την ακεραιότητα ή την ακρίβεια των πληροφοριών στον Ιστό. Από τυπική και νομική άποψη, η ακεραιότητα των πληροφοριών αφορά τη διασφάλιση από ακατάλληλη τροποποίηση ή καταστροφή των πληροφοριών και κυρίως, τη διασφάλιση της αυθεντικότητας των πληροφοριών.

Πράγματι, σε ιστότοπους όπως το Buy Likes and Followers, τα likes στο Facebook, οι ακόλουθοι στο Instagram, ακόμη και οι προβολές βίντεο στο YouTube έχουν όλα μια τιμή. Για παράδειγμα, στο Boostlikes 39 \$ δίνουν μια εφάπαξ παραγγελία 1.000 ακολούθων στο Instagram. Επιπλέον, σε μια μυστική έρευνα για το θέμα, οι Sunday Times μπόρεσαν να προσλάβουν επαγγελματίες συγγραφείς ψεύτικων κριτικών για μόλις 3 £ για κάθε κριτική πέντε αστέρων (Henry, 2015). Αυτές οι υπηρεσίες είναι διαθέσιμες σε οποιονδήποτε μπορεί και επιθυμεί να πληρώσει για αυτές, ανεξάρτητα από το ποιες είναι οι προθέσεις του αγοραστή ή οι συνέπειες που μπορεί να ακολουθήσουν. Αυτό καθιστά το crowdurfing ένα δυνητικά επιβλαβές φαινόμενο από διάφορες απόψεις. Πρώτα απ' όλα, οι μη αυθεντικές διαδικτυακές συμπεριφορές αποτελούν σημαντική απειλή για την αξιοπιστία των πληροφοριών που κυκλοφορούν στον Ιστό. Ειδικά οι διαδικτυακές κριτικές έχουν μεγάλη επιρροή στη διαμόρφωση των απόψεων των καταναλωτών και στις επακόλουθες αγοραστικές συμπεριφορές.

Μια έκθεση της Βρετανικής Αρχής Ανταγωνισμού και Μάρκετινγκ εκτιμά ότι ετησίως 23 δισεκατομμύρια £ των καταναλωτικών δαπανών του Ηνωμένου Βασιλείου επηρεάζονται από τις διαδικτυακές κριτικές. Δεδομένου ότι οι αξιόπιστες κριτικές είναι συχνά απαραίτητες για την αξιοπιστία ενός ιστότοπου λιανικής, ορισμένοι διαδικτυακοί έμποροι λιανικής, όπως η Amazon, έχουν οδηγηθεί σε μια φαινομενικά ατελείωτη μάχη για την εξάλειψη ιστοτόπων επί πληρωμή κριτικών που συνεχώς αναδύονται (Rubin, 2016). Δεύτερον, δεδομένου ότι ο καταγισμός με αναξιόπιστες πληροφορίες υπονομεύει την αξιοπιστία της πλατφόρμας, το crowdurfing μπορεί τελικά να παρεμποδίσει την αφομοίωση δυνητικά χρήσιμων πληροφοριακών συστημάτων (information systems – IS). Για παράδειγμα, μία έρευνα δείχνει ότι ο

πολλαπλασιασμός της ανήθικης συμπεριφοράς είναι ένας σημαντικός παράγοντας της μη αναστρέψιμης συμπεριφοράς διακοπής του IS από εκείνους που αισθάνονται εξαπατημένοι (Boukef & Charki, 2014).

Τρίτον, τέτοιες υπηρεσίες παρέχουν σε φορείς όπως πολιτικά κόμματα, κυβερνητικούς αξιωματούχους ή άλλους ισχυρά εργαλεία για την παραγωγή συναίνεσης που είναι σύμφωνη με την ατζέντα τους. Εάν οι φορείς με κακόβουλες προθέσεις είναι σε θέση να εκμεταλλευτούν πλήρως το δυναμικό αυτών των υπηρεσιών, η ζημιά μπορεί να μην περιορίζεται στην εξαπάτηση των καταναλωτών να αγοράσουν μη ποιοτικά προϊόντα: το crowdfunding μπορεί να αποτελέσει απειλή για κοινωνικές αρετές όπως η ελευθερία, η ακεραιότητα ή η σταθερότητα. Για παράδειγμα, αποκαλύφθηκε ότι η κινεζική κυβέρνηση έχει στρατολογήσει έναν στρατό από turfs – επίσης γνωστό ως «στρατός των 50 cents» (Han, 2015) – για να προωθήσει την επίσημη ατζέντα της στον εικονικό χώρο. Δεδομένου ότι η ανθρώπινη γνώση καταγράφεται και ανακτάται από το Διαδίκτυο σε αυξανόμενες ποσότητες, μια πρακτική όπως το crowdfunding μπορεί στην πιο ακραία μορφή της να διευκολύνει το σχηματισμό μιας Οργουελικής κοινωνίας όπου η γνώση ελέγχεται από εκείνους που έχουν αποκτήσει τα εργαλεία για τη διαμόρφωσή της με τις δικές τους ανάγκες.

Η βιομηχανία του crowdfunding υπολογίστηκε σε εκατομμύρια δολάρια το 2011, με προσδοκίες εκθετικής ανάπτυξης. Σχετικά με την ανήθικη χρήση IS γενικά, οι Chatterjee et al. (2015) έχουν αναγνωρίσει την έλλειψη έρευνας με ηθική προοπτική σχετικά με το γιατί τα άτομα εμπλέκονται σε τέτοια συμπεριφορά. Επιπλέον, υπογραμμίζουν την ανάγκη για περαιτέρω διερεύνηση του ρόλου της τεχνολογίας στη διευκόλυνση της ανήθικης συμπεριφοράς. Δεδομένου ότι το crowdfunding είναι μια μάλλον νέα μορφή ανήθικης χρήσης IS, δεν έχει διεξαχθεί ακόμη πολλή έρευνα για το θέμα. Ενώ έχουν γίνει κάποιες προσπάθειες για την παροχή εργαλείων για τον εντοπισμό δραστηριοτήτων crowdfunding, δεν γνωρίζουμε πολλά για τις συμπεριφορικές και ηθικές πτυχές του φαινομένου. Έτσι, ερωτήματα όπως τι παρακινεί τους εργαζόμενους στο crowdfunding, τους ενδιάμεσους πράκτορες και τους πελάτες, πώς δικαιολογούν τη συμπεριφορά στον εαυτό τους ώστε να μπορούν να λειτουργούν με καθαρή συνείδηση και πώς θα μπορούσαμε να αντιμετωπίσουμε αυτό το πρόβλημα, παραμένουν αναπάντητα.

Για να αντιμετωπίσουμε αυτή την προφανή ανάγκη για διερεύνηση των ηθικών και συμπεριφορικών πτυχών του crowdurfing, ξεκινήσαμε να βρούμε απαντήσεις στο ακόλουθο ερευνητικό ερώτημα: «Πώς οι διάφοροι παράγοντες που εμπλέκονται στο crowdurfing δικαιολογούν τη συμπεριφορά τους ηθικά;». Αντιμετωπίζουμε αυτό το ερώτημα διερευνώντας το φαινόμενο μέσα από θεωρητικούς φακούς που παρέχονται από την τρέχουσα βιβλιογραφία για την ασφάλεια στον κυβερνοχώρο και την ηθική στη χρήση IS. Συγκεκριμένα, υποστηρίζουμε ότι η θεωρία της αποτροπής, το πιο ευρέως αναφερόμενο πλαίσιο στην έρευνα για την ασφάλεια στον κυβερνοχώρο, μπορεί να μην είναι αποτελεσματική στην αποτροπή του crowdurfing. Έτσι, εξετάζουμε γεγονότα που προηγούνται των προθέσεων για κατάχρηση συστήματος (Willison & Warkentin, 2013), αλλά και το μετα-πλαίσιο ηθικής για συστήματα πληροφοριών του Smith (2002). Παράλληλα, πραγματοποιούμε μια εννοιολογική ανάλυση των ηθικών λογικών πίσω από το crowdurfing, εφαρμόζοντας το πλαίσιο σε μια πραγματική επιχείρηση crowdurfing - το Paid Book Reviews, ένας ιστότοπος που υπόσχεται στους πελάτες του που πληρώνουν τη διευθέτηση θετικών κριτικών για αντικείμενα που πωλούνται στο Amazon.

Το crowdurfing αντιπροσωπεύει μια από τις πολλές μορφές που έχει λάβει το crowdsourcing την τελευταία δεκαετία. Τυπικά, θεωρούνται μια ηλεκτρονική αγορά, όπου οι αναζητητές (συνήθως εταιρείες που αναζητούν λύσεις σε συγκεκριμένα προβλήματα) και οι λύτες (δυναμικά οποιοσδήποτε στο πλήθος) είναι στο εμπόριο λύσεων για ανταμοιβές μέσω της πλατφόρμας crowdsourcing (συχνά λειτουργεί από τρίτο). Ωστόσο, η διαφορά μεταξύ του crowdsourcing και του crowdurfing συνοψίζεται σε ένα μόνο χαρακτηριστικό: τη φύση των διαπραγματεύσιμων εργασιών. Από τη μία πλευρά, σε γενικό επίπεδο, το crowdsourcing έχει επαινεθεί τόσο από ακαδημαϊκούς όσο και από επαγγελματίες ως επιθυμητό φαινόμενο και κυρίως λόγω των πολλών κοινωνικών αξιών που προέκυψαν από τη βελτιστοποίηση των «υποχρησιμοποιημένων πόρων» του πλήθους, π.χ., η συμμετοχή στη δημοσιογραφία (Soliman, 2015).

Από την άλλη πλευρά, το crowdurfing είναι μια συγκεκριμένη περίπτωση crowdsourcing στην οποία τα μέλη του πλήθους πληρώνονται για να αφήσουν ένα ψεύτικο ψηφιακό ίχνος, π.χ., προσποιούμενοι ότι είναι θαυμαστές κάποιου

λογαριασμού στο Instagram. Ενώ το crowdsourcing από την άποψη της δημιουργίας αξίας απολαμβάνει αυξανόμενο ενδιαφέρον ως τρέχον θέμα έρευνας του IS, έχει δοθεί λιγότερη προσοχή στη σκοτεινή πλευρά του crowdsourcing που συχνά οδηγεί σε υποβάθμιση της αξίας των πληροφοριών στο Διαδίκτυο, καθώς ο θεμελιώδης στόχος τείνει να είναι για την προώθηση των συμφερόντων ορισμένων μερών σε βάρος του υπόλοιπου πληθυσμού του Διαδικτύου (Price, 2014).

Μια εκστρατεία crowdurfing περιλαμβάνει συνήθως τρεις βασικούς παράγοντες: εργαζόμενους που εκτελούν μια συγκεκριμένη εργασία έναντι αμοιβής, πελάτες που επωφελούνται και πληρώνουν για την εργασία και τον πράκτορα, συχνά σε μορφή ενδιάμεσης πλατφόρμας, που συνδέει τους εργαζόμενους με τους πελάτες και ενορχηστρώνει ολόκληρη τη συναλλαγή (Wang et al., 2012). Το προβλέψιμο αποτέλεσμα τέτοιων εκστρατειών σε μια πλατφόρμα-στόχο είναι η διαστρέβλωση της κοινής γνώμης μεταξύ των χρηστών της. Για αρκετά χρόνια, η Amazon ήταν ο στόχος τέτοιων καμπανιών όπου ένας αυτοδημοσιευμένος συγγραφέας μπορούσε να αγοράσει μια κριτική «πέντε αστέρων» για μόλις πέντε δολάρια (Shaffer, 2013).

Άλλες επιχειρήσεις που δραστηριοποιούνται σε αυτόν τον τομέα διευκολύνουν την αγορά των «likes» στο Facebook, την παρακολούθηση λογαριασμών μέσω κοινωνικής δικτύωσης, το retweet ορισμένων μηνυμάτων στο twitter, ακόμη και την εκτέλεση ακριβών αναζητήσεων στο Google και κάνοντας κλικ σε έναν καθορισμένο σύνδεσμο για να ανεβάσουν την κατάταξή του. Επιπλέον, οι εκστρατείες crowdurfing επηρεάζουν και άλλους ενδιαφερόμενους στο Διαδίκτυο. Τέτοιες εκστρατείες αποτελούν απειλή για την πλατφόρμα-στόχο όπου οι εργαζόμενοι ανατίθενται να εκτελέσουν τα καθήκοντά τους (π.χ. Amazon, Facebook, Twitter, Google), καθώς η αύξηση των πλαστών πληροφοριών μπορεί να επιδεινώσει την αξιοπιστία και τη χρησιμότητα αυτής της πλατφόρμας. Επιπλέον, όταν κοιτάζουμε πέρα από αυτά τα τέσσερα άμεσα επηρεαζόμενα μέρη, το crowdurfing επηρεάζει ολόκληρη την κοινότητα του Διαδικτύου (κάτι που μπορεί τελικά να επηρεάσει τις κοινότητες στο φυσικό πεδίο), καθώς οι εκστρατείες crowdurfing διαμορφώνουν τη φύση και την ποιότητα των πληροφοριών που είναι διαθέσιμες στο Διαδίκτυο (Wang et al., 2012).

Χρησιμοποιώντας πραγματικούς ανθρώπους εργαζομένους, οι πελάτες αντιμετωπίζουν τις αδυναμίες της χρήσης bots για τη δημιουργία λογαριασμών και περιεχομένου που

δημιουργούνται αυτόματα. Αν και έχουν αναπτυχθεί μέσα για την αναστολή των κακόβουλων εργασιών που εκτελούνται αυτόματα από τα ρομπότ), θεωρούνται λιγότερο αποτελεσματικά έναντι των ανθρώπων που μπορούν να φαίνονται σχετικά ειλικρινείς και γνήσιοι. Μέχρι στιγμής, το crowdurfing έχει αντιμετωπιστεί σχεδόν αποκλειστικά στον τομέα της επιστήμης των υπολογιστών, κυρίως με την ανάπτυξη αλγορίθμων και εργαλείων μηχανικής μάθησης που μπορούν να ανιχνεύσουν και να αναγνωρίσουν τους crowdurfers και τις κακόβουλες δραστηριότητές τους. Οι συμβατικές μέθοδοι ανίχνευσης βασίζουν συνήθως τις αναλύσεις τους σε χαρακτηριστικά λογαριασμού και έχουν καταφέρει να παράγουν ποσοστά ανίχνευσης έως και 97,35% για εργασίες crowdurfing (Lee et al., 2015).

Ωστόσο, μια βασική πρόκληση για την ανίχνευση του crowdurfing είναι ότι εφόσον οι turfers είναι πραγματικοί εργαζόμενοι, τα χαρακτηριστικά του λογαριασμού τους μπορεί να είναι πολύ παρόμοια με τους κανονικούς, μη κακόβουλους λογαριασμούς. Έτσι, έχουν εισαχθεί και σύγχρονες μέθοδοι, όπως το CrowdTarget που αντί για λογαριασμούς και δραστηριότητες turfer εντοπίζει τα ίδια τα αντικείμενα-στόχους που βρίσκονται υπό απόπειρες χειραγώγησης (Song et al., 2015).

Ενώ η εργασία για την ανίχνευση του crowdurfing έχει συμβάλει ανεκτίμητα στην κατανόηση του προβλήματος και προσφέρει πρακτική βοήθεια στην καταπολέμησή του, η προσέγγιση αφήνει έναν σημαντικό τομέα ανεξήγητο: τη συμπεριφορά των χρηστών. Για να εξηγηθεί πλήρως και να αντιμετωπιστεί αποτελεσματικά ένα τέτοιο φαινόμενο, θα πρέπει κανείς να κατανοήσει τους ηθικούς λόγους πίσω από τη συμπεριφορά. Παραδόξως, μέχρι σήμερα δεν υπάρχει καμία έρευνα που να στοχεύει να εξηγήσει τα κίνητρα και τις δικαιολογίες πίσω από το crowdurfing. Επιπλέον, ενώ η προηγούμενη βιβλιογραφία έχει καταφέρει να ανιχνεύσει το crowdurfing, εξακολουθούμε να μην έχουμε τα εργαλεία για την καταπολέμηση και την αποτροπή του. Για να αντιμετωπίσουμε αυτό το φαινομενικό ερευνητικό κενό, συζητάμε πρώτα πώς αυτά τα ζητήματα θα μπορούσαν να αντιμετωπιστούν με τις υπάρχουσες θεωρίες για την ασφάλεια των πληροφοριών και, στη συνέχεια, αναζητούμε εξηγήσεις για το crowdurfing βλέποντάς το μέσα από διαφορετικές προοπτικές κανονιστικής ηθικής (Rinta-Kahila & Soliman, 2017).

3.2 Η σκοτεινή πλευρά του crowdsourcing

Μέχρι στιγμής, έχουμε υιοθετήσει μια κανονιστική θέση που χαρακτηρίζει το crowdsurfing ως τη σκοτεινή πλευρά του crowdsourcing, δηλαδή μια ανήθικη συμπεριφορά στον κυβερνοχώρο που αξίζει να διερευνηθεί πώς να την αναχαιτίσουμε ή να την αποτρέψουμε. Στο ευρύ φάσμα της συμπεριφορικής βιβλιογραφίας για την ασφάλεια πληροφοριών, η γενική θεωρία αποτροπής από την εγκληματολογία ήταν μια από τις πιο αναφερόμενες θεωρίες που υποθέτει ότι οι παραβατικές συμπεριφορές αποτρέπονται μέσω της εφαρμογής αυστηρών και ορισμένων ποινών. Για παράδειγμα, οι Straub & Nance (1990) χρησιμοποιούν τη θεωρία της αποτροπής ως θεμέλιο του συνολικού πλαισίου τους για τη μελέτη των κινδύνων ασφάλειας του IS. Στον κύκλο ενεργειών ασφαλείας τους, η διαδικασία μείωσης των κινδύνων ασφαλείας των συστημάτων κορυφώνεται σε τέσσερις διαδοχικές δραστηριότητες: 1) αποτροπή, 2) πρόληψη, 3) ανίχνευση και 4) διορθωτικά μέτρα. Εν ολίγοις, αυτή η θεωρία υποστηρίζει ότι η αντιμετώπιση των κινδύνων ξεκινά από την αποτροπή που τονίζει την κοινοποίηση της βεβαιότητας και της αυστηρότητας της επιβολής κυρώσεων στον θύτη. Ωστόσο, εάν αυτό δεν καταφέρει να αποτρέψει τον καταχραστή, το επόμενο βήμα είναι η αποτροπή της απειλής με ενεργά αντίμετρα. Εάν επίσης αποτύχει η πρόληψη, ο κάτοχος του συστήματος θα πρέπει να είναι σε θέση να ανιχνεύσει την απειλή. Τέλος, θα πρέπει να ληφθούν μέτρα για την αποκατάσταση των ζημιών και την τιμωρία του θύτη.

Η αποτελεσματική δράση σε κάθε στάδιο ενισχύει την αποτροπή καθώς οι απαντήσεις σε παραβιάσεις σε κάθε στάδιο ενημερώνουν τους πιθανούς καταχραστές για τις συνέπειες της κατάχρησης (δηλαδή, βρόχος ανατροφοδότησης αποτροπής). Ωστόσο, η αποτελεσματικότητα των κυρώσεων (που έχουν τις ρίζες τους στο θεωρητικό πλαίσιο της αποτροπής) υπονομεύεται σε περιβάλλοντα που στερούνται ισχυρής κεντρικής διακυβέρνησης, καταναγκασμού και «μαζικής υπακοής» και όταν οι φορείς χρησιμοποιούν τακτικές εξουδετέρωσης που τους επιτρέπουν να ελαχιστοποιούν την αντιληπτή βλάβη της συμπεριφοράς τους (Siponen et al., 2012).

Όσον αφορά τα προληπτικά μέτρα, οι Levchenko et al. (2011) προτείνουν ότι η παρακολούθηση της ροής χρημάτων θα μπορούσε να είναι αποτελεσματική για την καταπολέμηση του ανεπιθύμητου περιεχομένου. Διαπιστώνουν ότι η χρηματοδότηση για το 95% των φαρμακευτικών προϊόντων, των αντιγράφων και των προϊόντων λογισμικού που διαφημίζονται σε ανεπιθύμητα μηνύματα περνάει από λίγες μόνο τράπεζες. Έτσι, προτείνουν ότι να πειστούν οι τράπεζες να αρνηθούν την παροχή ορισμένων υπηρεσιών σε γνωστούς αποστολείς ανεπιθύμητης αλληλογραφίας θα μπορούσε να λειτουργήσει ως αποτελεσματική πολιτική παρέμβασης που θα αποτρέψει την αποστολή ανεπιθύμητης αλληλογραφίας. Ωστόσο, ενώ η παρακολούθηση της ροής των χρημάτων θα μπορούσε να βοηθήσει στην αναχαίτιση του crowdfunding, το γεγονός ότι αρκετές πλατφόρμες πρακτόρων δέχονται κρυπτογραφημένες πληρωμές bitcoin το καθιστά όλο και πιο περίπλοκο. Επιπλέον, η υπάρχουσα έρευνα για το crowdfunding ασχολείται συγκεκριμένα με το τρίτο στάδιο του κύκλου: την ανίχνευση.

Η αποτελεσματική ανίχνευση του crowdfunding επέτρεψε στις πλατφόρμες-στόχους να εφαρμόσουν διορθωτικές ενέργειες, όπως η απαγόρευση ορισμένων πωλητών και αναθεωρητών από τη χρήση των υπηρεσιών τους και η άσκηση νομικής αγωγής κατά των πλατφορμών πρακτόρων (Rubin, 2016). Αλλά αυτό συμβαίνει συχνά με καθυστέρηση και οι κακόβουλοι δρώντες μπορούν πάντα να δημιουργήσουν νέους λογαριασμούς για να συνεχίσουν την κατάχρηση. Αναγνωρίζοντας αυτούς τους περιορισμούς, υπήρξαν πρόσφατες εκκλήσεις να επεκταθεί το ερευνητικό πεδίο για να υπερβεί την αποτροπή και να διερευνήσει τα «προ-κινητικά» γεγονότα ή φαινόμενα που συμβαίνουν προσωρινά πριν από την αποτροπή (Willison & Warkentin, 2013). Τέτοια γεγονότα προάγουν τον σχηματισμό προθέσεων συμπεριφοράς για κατάχρηση ενός συστήματος, που προκύπτουν από την αλληλεπίδραση μεταξύ του χρήστη IS και του περιβάλλοντος (π.χ. οργανισμός) και εμφανίζονται με μορφές όπως αίσθημα αδικίας, δυσαρέσκεια, εξουδετέρωση και αυτοέκφραση.

Για παράδειγμα, οι Willison & Warkentin (2013) προσδιορίζουν τα εργαλειακά και εκφραστικά κίνητρα ως εξέχοντες οδηγούς για την κατάχρηση συστήματος. Ένα άλλο προ-κινητικό γεγονός που προωθούμε εδώ είναι η σημασία της κατανόησης των διαφόρων κανονιστικών (ηθικών) επιχειρημάτων που μπορούν να εφαρμόσουν όσοι

εμπλέκονται στο crowdurfing για να δώσουν στη συμπεριφορά τους ηθική νομιμότητα. Η ακόλουθη ανάλυση επηρεάζεται σε μεγάλο βαθμό από την πραγματεία του Smith (2002) για την ηθική στα πληροφοριακά συστήματα.

3.3 Ηθικά ζητήματα στη χρήση IS

Οι ανησυχίες για την ηθική συμπεριφορά των ανθρώπων στην εποχή της πληροφορίας κάθε άλλο παρά καινούργιο είναι. Ο Mason (1986) αποδίδει αυτή την ανησυχία σε ένα θεμελιώδες χαρακτηριστικό της πληροφορίας: «η πληροφορία είναι το μέσο μέσω του οποίου ο νους επεκτείνεται και αυξάνει την ικανότητά του να επιτύχει τους στόχους του, συχνά ως αποτέλεσμα μιας εισροής από άλλο μυαλό» (Mason, 1986). Με βάση αυτή την υπόθεση, ο Mason (1986) ζητά μια αναθεώρηση του κοινωνικού συμβολαίου που διέπει τους ανθρώπους στην εποχή της πληροφορίας, με ιδιαίτερη προσοχή σε τέσσερα ηθικά ζητήματα που αντικατοπτρίζουν τέσσερα τρωτά σημεία: (1) ιδιωτικότητα, (2) ακρίβεια, (3) ιδιοκτησία και (4) προσβασιμότητα.

Το απόρρητο είναι θέμα έκθεσης και αφορά ποιες πληροφορίες πρέπει να αποκαλύψει ή/και να κρατήσει μυστικό ένα άτομο. Η ακρίβεια είναι ζήτημα αυθεντικότητας και ασχολείται με το πώς να διασφαλιστεί ότι οι υπεύθυνοι λήψης αποφάσεων (όποιοι κι αν είναι) δεν είναι παραπληροφορημένοι. Η ιδιοκτησία αφορά ποιος θα πρέπει να κατέχει την πνευματική ιδιοκτησία των πληροφοριών και με ποιο τίμημα. Η προσβασιμότητα είναι ζήτημα αποδοχής και αφορά τη διασφάλιση της παραχώρησης δικαιωμάτων πρόσβασης στις πληροφορίες μόνο στα επιδιωκόμενα άτομα. Αν και το crowdurfing μπορεί να μην απειλεί το απόρρητο, την ιδιοκτησία ή την προσβασιμότητα, αποτελεί αναμφισβήτητα σημαντική απειλή για την ακρίβεια των πληροφοριών (Rinta-Kahila & Soliman, 2017).

3.3.1 Παραδοσιακή Φιλοσοφική Ηθική

Με βάση το έργο του Mason (1986), οι Chatterjee et al. (2015) ορίζουν την ανήθικη χρήση της πληροφορικής ως «εσκεμμένη παραβίαση—από οποιοδήποτε άτομο, ομάδα ή οργανισμό—του απορρήτου ή/και περιουσίας ή/και ακρίβειας και/ή πρόσβασης—σε

Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ

σχέση με πληροφορίες/πληροφοριακά αγαθά ή μέρος ενός πληροφοριακού συστήματος, που ανήκει/ ελέγχεται από οποιοδήποτε άλλο άτομο, ομάδα ή οργανισμό». Εγγενής σε αυτόν τον ορισμό είναι η υπόθεση ότι μια παραβίαση είναι ανήθικη εάν προκαλεί βλάβη ή/και παραβιάζει τους κωδικοποιημένους κανόνες, αντιπροσωπεύοντας δύο ανταγωνιστικές σχολές ηθικής σκέψης: την τελεολογική και τη δεοντολογική σχολή. Ο Smith (2002) σημειώνει ότι η βασική ένταση μεταξύ αυτών των δύο παραδειγμάτων συνοψίζεται στην αντίληψη του «σωστού» ή του «λάθους». Η τελεολογική προοπτική (γνωστή και ως συνεπακόλουθες θεωρίες), από τη μια πλευρά, υποθέτει ότι «η σωστή απάντηση σε ένα ηθικό δίλημμα εξαρτάται από τις συνέπειες που προκύπτουν από τη δράση - κάποιος πρέπει να επιλέξει την εναλλακτική με συνέπειες που παράγουν το πιο «καλό» για κάποια ομάδα.

Από την άλλη πλευρά, η δεοντολογική προοπτική, υποστηρίζει ότι το «ηθικό», «σωστό» ή «λάθος» μιας πράξης δεν μπορεί ποτέ να προσδιοριστεί αποκλειστικά από τις συνέπειες που απορρέουν από αυτήν. Αντίθετα, το τι είναι «σωστό» καθορίζεται από ορισμένες γενικές αρχές. Συχνά, η εστίαση είναι στην πρόθεση πίσω από μια ενέργεια. Ακόμα κι αν τα πράγματα πηγαιναν άσχημα, ένας κατηγορηματικός στοχαστής θα μπορούσε να ισχυριστεί ότι μια ενέργεια ήταν ηθικά σωστή, αν η υποκείμενη πρόθεση του δράντος ήταν σωστή». Με βάση αυτή την άποψη, ο Kant και άλλοι έχουν προτείνει γενικές κατευθυντήριες γραμμές όπως το «δεν πρέπει ποτέ να λες ψέματα» (Smith 2002).

3.3.2 Επιχειρηματική Ηθική

Εκτός από τις παραδοσιακές φιλοσοφικές θεωρίες, ο Smith (2002) συζητά την επιχειρηματική ηθική, παρέχοντας τις δικές τους κατευθυντήριες γραμμές για το τι είναι σωστό και τι είναι λάθος. Πρώτον, η θεωρία των μετόχων θεωρεί ότι οι υπεύθυνοι λήψης αποφάσεων στις εταιρείες πρέπει να λαμβάνουν τέτοιες ενέργειες που μεγιστοποιούν τα κέρδη και δημιουργούν τη μεγαλύτερη αξία για τους μετόχους μακροπρόθεσμα. Αντίθετα, η θεωρία των ενδιαφερομένων υποστηρίζει ότι «τα συμφέροντα των ενδιαφερομένων εκτός των μετόχων θα πρέπει να λαμβάνονται υπόψη μαζί με αυτά των μετόχων, ακόμη κι αν αυτό μειώνει την κερδοφορία της επιχείρησης» (Smith 2002).

Επίσης, ο Smith (2002) συζητά τη θεωρία των κοινωνικών συμβολαίων της επιχειρηματικής ηθικής, η οποία «απαιτεί από τους διευθυντές να ενεργούν έτσι ώστε τα συμφέροντα των καταναλωτών και των εργαζομένων να ικανοποιούνται με τρόπο που μεγιστοποιεί τα πλεονεκτήματα και ελαχιστοποιεί τα μειονεκτήματα». Επιπλέον, η θεωρία του κοινωνικού συμβολαίου δίνει εξαιρετική έμφαση στην έννοια της δικαιοσύνης. Για παράδειγμα, σε ένα πλαίσιο επιχειρηματικής ηθικής, ο όρος δικαιοσύνη ενός υποθετικού κοινωνικού συμβολαίου ορίζει ότι οι εταιρείες πρέπει να αποφεύγουν την «απάτη», την «εξαπάτηση», καθώς και κάθε πρακτική που «επιδεινώνει συστηματικά την κατάσταση μιας δεδομένης ομάδας στην κοινωνία» (Smith, 2002).

Ομοίως, για ένα άτομο, η προσκόλληση στο κοινωνικό συμβόλαιο μπορεί να παρουσιαστεί π.χ. ως υπακοή στο νόμο και συμπεριφορά που δεν προκαλεί βλάβη στους άλλους - θα απέφευγε κανείς την κλοπή από άλλους ή την κακοποίηση άλλων, παρόλο που η κακόβουλη συμπεριφορά θα είχε ορισμένα οφέλη για το άτομο που πράττει. Η θεωρία των κοινωνικών συμβολαίων είναι χρήσιμη στη διερεύνηση ορισμένων αμφιλεγόμενων συμπεριφορών της εποχής της πληροφορίας, π.χ. Διαφήμιση στο Διαδίκτυο (Gordon & De Lima-Turner, 1997).

3.3.3 Κωδικοποιημένοι κανόνες

Επιπροσθέτως, ο Smith (2002) εξηγεί ότι οι κωδικοποιημένοι κανόνες μπορούν να χρησιμοποιηθούν για να παρακάμψουν την ανάγκη για κανονιστική επιχειρηματολογία που είναι εγγενής στην εφαρμογή των σχολών ηθικής που συζητήθηκαν προηγουμένως. Στην ουσία, οι κωδικοποιημένοι κανόνες είναι «ένα σύνολο κατευθυντήριων γραμμών που μπορούν να ακολουθηθούν» και οι οποίες «μπορούν να δημιουργηθούν σε διάφορα επίπεδα, κυρίως για μια συγκεκριμένη εταιρεία (συντάσσεται «Εταιρικός Κώδικας Δεοντολογίας», η παραβίαση του οποίου είναι μερικές φορές αιτία απόλυσης), για ένα επάγγελμα ή κλάδο (συντάσσεται ως «Κώδικας Δεοντολογίας»), ή για μια κοινωνία στο σύνολό της, όπως μαρτυρείται στη νομική της δομή» (Smith 2002). Είναι ενδιαφέρον ότι το πλαίσιο IS επιτρέπει τη σχετικά εύκολη εφαρμογή πολύ συγκεκριμένων και λεπτομερών επιπέδων κωδικοποιημένων κανόνων, όπως οι όροι και οι προϋποθέσεις συγκεκριμένης χρήσης

λογισμικού ή διαφορετικά σύνολα κανόνων για χρήστες στους οποίους παραχωρούνται διαφορετικά επίπεδα δικαιωμάτων τροποποίησης σε έναν ιστότοπο.

3.4 Η σύγκρουση του κυβερνοχώρου και του φυσικού κόσμου

Εν ολίγοις, οι θεωρίες που συζητήθηκαν παραπάνω μπορούν να βοηθήσουν στην κατανόηση της ανήθικης χρήσης του διαδικτύου, ειδικά αν λάβουμε υπόψη τη διάκριση μεταξύ εικονικής και φυσικής πραγματικότητας. Παρόλο που ο εικονικός κόσμος συγχωνεύεται ολοένα και περισσότερο με το φυσικό πεδίο (π.χ., η εμφάνιση της εικονικής πραγματικότητας και των εφαρμογών επαυξημένης πραγματικότητας όπως το Pokémon Go), από την άποψη της συμπεριφοράς, αυτοί οι δύο χώροι εξακολουθούν να θεωρούνται σε μεγάλο βαθμό διακριτοί μεταξύ τους. Για παράδειγμα, ένα άτομο που θα θεωρούσε την κλοπή μιας ταινίας ή ενός δίσκου μουσικής από ένα κατάστημα λιανικής ως εξαιρετικά ανήθικο, μπορεί να είναι απολύτως εντάξει με την παράνομη λήψη του ίδιου προϊόντος (Siponen, 2004).

Συνολικά, φαίνεται ότι η συμπεριφορά των ατόμων έχει λιγότερες αναστολές στον κυβερνοχώρο από ό,τι στον φυσικό κόσμο. Μια εξήγηση σε αυτό θα μπορούσε να είναι η πραγματική ή η αντιληπτή ανωνυμία στο Διαδίκτυο που μειώνει τη λογοδοσία. Αυτό υποδηλώνει ότι ένας ορισμένος βαθμός ανωνυμίας και η έλλειψη φυσικής παρουσίας που επιτρέπονται από τον κυβερνοχώρο θα μπορούσαν πραγματικά να βοηθήσουν στην ανάδειξη του χειρότερου εαυτού στους ανθρώπους. Ο Leetaru (2016) υποστηρίζει ότι το Διαδίκτυο μπορεί να εξελίσσεται μακριά από τον αρχικό του ρόλο ως το λίκνο της ελευθερίας του λόγου κατά της λογοκρισίας και προς ένα μέσο ελεγχόμενο από τις εταιρείες που παρακολουθεί τα συμφέροντα του εμπορίου και των κυβερνήσεων εναντίον των οποίων επαναστατούσε. Έτσι, οι τρέχουσες συγκρούσεις και τα προβλήματα που σχετίζονται με ζητήματα όπως η εικονική κατάχρηση, τα άυλα δικαιώματα ιδιοκτησίας, η αιτιολόγηση της επιτήρησης από τους χρήστες του Διαδικτύου και άλλα θα μπορούσαν να σχετίζονται με αυξανόμενες προκλήσεις στη σύγκρουση εικονικής και φυσικής πραγματικότητας πριν βρεθεί μια νέα ισορροπία. Είναι προφανές ότι ο κοινωνικός ρόλος του κυβερνοχώρου εξακολουθεί να εξελίσσεται: ενώ οι άνθρωποι κάνουν τις δικές τους ερμηνείες σχετικά με το τι είναι

Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ

Διπλωματική Εργασία: Crowdfunding και E-commerce: Επισκόπηση και μελέτη των επιπτώσεων σε συστήματα ηλεκτρονικού εμπορίου

αποδεκτό στον κυβερνοχώρο, οι οργανισμοί αγωνίζονται να συμβαδίσουν με την αυξανόμενη ανάγκη για νέες νομοθεσίες και ηθικές κατευθυντήριες γραμμές. Ως εκ τούτου, μπορεί να υπάρχει μια υστέρηση μεταξύ της συμμόρφωσης των ανθρώπων σε ηθικές κατευθυντήριες γραμμές, όπως κοινωνικά συμβόλαια ή κωδικοποιημένους κανόνες στον κυβερνοχώρο για συμπεριφορές που κατά τα άλλα είναι απαράδεκτες στον φυσικό χώρο (Leetaru, 2016).

ΚΕΦΑΛΑΙΟ 4ο

4.1 Crowdturfing και Social Media

Εάν οι συμμετέχοντες σε μια εκστρατεία astroturfing είναι οργανωμένοι μέσω crowdsourcing, η διαδικασία ορίζεται ως Crowdturfing. Το crowdturfing στοχεύει να κερδίσει ή να καταστρέψει τη φήμη ανθρώπων, προϊόντων και άλλων οντοτήτων μέσω της διάδοσης μεροληπτικών απόψεων και πλαισιωμένων πληροφοριών. Το Crowdturfing Detection είναι η διαδικασία ανίχνευσης δραστηριοτήτων crowdturfing με συστήματα λογισμικού. Η μέθοδος ανίχνευσης μπορεί να είναι χωρίς επίβλεψη μέσω της ανακάλυψης της ανώμαλης κυκλοφορίας ή βάσει προτύπων ιστορικών δραστηριοτήτων crowdturfing. Το Crowdturfing στα Social Media είναι οι δραστηριότητες crowdturfing που θεωρούν τις πλατφόρμες κοινωνικής δικτύωσης ως το κύριο κανάλι ενημέρωσης της καμπάνιας. Οι εργαζόμενοι στο crowdturfing χρησιμοποιούν λογαριασμούς μέσω κοινωνικής δικτύωσης για να διαδώσουν πληροφορίες – γεγονός που μπορεί να οδηγήσει σε άδικη δημοτικότητα (Wu & Liu, 2018).

Η διάχυτη χρήση των υπηρεσιών κοινωνικής δικτύωσης τα τελευταία χρόνια έχει φέρει επανάσταση στον τρόπο διάδοσης της πληροφορίας. Οι χρήστες των υπηρεσιών μέσω κοινωνικής δικτύωσης έχουν γίνει όχι μόνο καταναλωτές πληροφοριών, αλλά και παραγωγοί πληροφοριών. Το άνοιγμα και η διαθεσιμότητα των υπηρεσιών κοινωνικής δικτύωσης διευκολύνουν επίσης τους κακόβουλους χρήστες να κάνουν κακή χρήση των υπηρεσιών κοινωνικής δικτύωσης. Το crowdturfing αντικατοπτρίζεται σε τέτοιες κακόβουλες δραστηριότητες, οι οποίες εκτελούνται από πραγματικούς ανθρώπους και οργανώνονται από crowdsourcing. Με την αυξανόμενη διαθεσιμότητα των μέσω κοινωνικής δικτύωσης, οι άνθρωποι ακούν και εμπιστεύονται τις απόψεις των διαδικτυακών τους φίλων προτού λάβουν αποφάσεις. Επίσης, χρησιμοποιούν πληροφορίες από τα κοινωνικά δίκτυα για να βρουν ενδιαφέρουσες ταινίες, εστιατόρια, κ.λπ. Ωστόσο, αυτές οι θετικές εμπειρίες παρουσιάζουν επίσης ένα δυσοίωνα κομμάτι,

καθώς ψεύτικες και ανακριβείς πληροφορίες διαδίδονται σκόπιμα για να παραπλανήσουν τους ανθρώπους (Wu & Liu, 2018).

Τις περισσότερες φορές, οι άνθρωποι υπέθεταν ότι οι κακόβουλες δραστηριότητες δημιουργούνταν αυτόματα από αυτοματοποιημένα συστήματα, επομένως τα υπάρχοντα συστήματα που εξαρτώνται από την υπόθεση είναι εύκολο να παρακαμφθούν από πραγματικούς χρήστες. Το crowdsourcing διευκολύνει τις επιθέσεις μέσω της συγκέντρωσης εργαζομένων στο crowdurfing και της σύνδεσης τους με πιθανούς πελάτες. Οι περίπλοκες επιθέσεις crowdurfing εκφοβίζουν τους απλούς χρήστες με συντριπτικές ανεπιθύμητες πληροφορίες. Ως εκ τούτου, είναι σημαντικό να ανιχνεύεται το crowdurfing στα μέσα κοινωνικής δικτύωσης (Wu & Liu, 2018).

Πριν προχωρήσουμε περαιτέρω, είναι σημαντικό να αναλύσουμε την έννοια του crowdsourcing και πώς το crowdsourcing οργανώνει το crowdurfing. Οι ιστότοποι crowdsourcing, όπως το Amazon Mechanical Turks, το Clickworker και το CrowdFlower, δίνουν τη δυνατότητα στους πελάτες να δημοσιεύουν θέσεις εργασίας και να προσλαμβάνουν εργαζόμενους στο crowdsourcing για να αποκτήσουν τις απαραίτητες υπηρεσίες. Οι υπηρεσίες μπορεί να είναι ποικίλες, που κυμαίνονται από απαντήσεις σε έρευνες, σχολιασμό εικόνων έως μετάφραση κειμένων. Δεδομένου ότι οι αποκλειστικοί σχολιαστές είναι συνήθως ακριβοί στην πρόσληψη, είναι δύσκολο να χρησιμοποιηθούν σχολιασμοί ή έρευνες μεγάλης κλίμακας. Το crowdsourcing μειώνει το κόστος τμηματοποιώντας μια μεγάλη εργασία σε μικρά κομμάτια και συνδέοντάς τα με τον ελεύθερο χρόνο των εργαζομένων στο crowdsourcing. Εν τω μεταξύ, ωστόσο, το crowdsourcing ανοίγει επίσης το δρόμο για κακόβουλες εκστρατείες μεγάλης κλίμακας (Wu & Liu, 2018).

Το Zhubajie.com (ZBJ) είναι ένας από τους μεγαλύτερους ιστότοπους crowdsourcing στην Κίνα, ο οποίος ξεκίνησε το 2006. Στο ZBJ, οι εργασίες ταξινομούνται σε διαφορετικές κατηγορίες, όπως προγραμματισμός (π.χ. δημιουργία ιστοσελίδων) και γραφιστική. Υπάρχει επίσης μια υποενότητα αφιερωμένη για αιτήματα αποκλειστικά για crowdurfing. Οι εργασίες του crowdurfing αφορούν κυρίως τη διαφήμιση ενός συγκεκριμένου προϊόντος: οι πελάτες συνήθως ζητούν από τους εργαζόμενους στο crowdurfing να δημοσιεύσουν ένα συγκεκριμένο περιεχόμενο και να επιστρέψουν τα

αντίστοιχα στιγμιότυπα οθόνης για να πληρωθούν. Ο μέσος προϋπολογισμός μιας εργασίας crowdurfing είναι περίπου 12 έως 15 δολάρια και ο κάθε εργαζόμενος στο crowdurfing μπορεί να κερδίσει 0,2 έως 1 δολάριο κατά μέσο όρο για μία μόνο εργασία. Το crowdurfing στα μέσα κοινωνικής δικτύωσης συνήθως εμπλέκεται με τη διάδοση κακόβουλων URL και τη δημιουργία καμπανιών astroturf. Ο εντοπισμός του crowdurfing στα μέσα κοινωνικής δικτύωσης θα μπορούσε να είναι επωφελής για τη βελτίωση των εμπειριών των χρηστών και τη διατήρηση της αξίας των ιστότοπων κοινωνικής δικτύωσης (Wang et al., 2012).

4.2 Η διαδικασία ανίχνευσης

Η μελέτη της ανίχνευσης του crowdurfing υπόκειται σε διάφορες προκλήσεις. Η διαθεσιμότητα ιστοσελίδων crowdsourcing και το άνοιγμα των μέσων κοινωνικής δικτύωσης κάνουν το πρόβλημα ακόμη πιο περίπλοκο. Παρακάτω παρουσιάζονται σημαντικές προκλήσεις που σημειώθηκαν από ερευνητές και επαγγελματίες (Karlan & Haenlein, 2010):

- **Απουσία βασικής αλήθειας:** Σε αντίθεση με τα παραδοσιακά προβλήματα εξόρυξης δεδομένων, η βασική αλήθεια των δεδομένων crowdurfing στα μέσα κοινωνικής δικτύωσης δύσκολα μπορεί να σχολιαστεί με μη αυτόματο τρόπο. Αν και είναι δυνατό για τους σχολιαστές να εντοπίσουν ακανόνιστα μοτίβα και δυσμενείς επιθέσεις, ο τρόπος οργάνωσης τέτοιων εκστρατειών, ωστόσο, δύσκολα μπορεί να εντοπιστεί με βεβαιότητα. Σύμφωνα με τον ορισμό του crowdurfing, η οργάνωση μέσω crowdsourcing είναι βασικό στοιχείο του crowdurfing. Επομένως, υπάρχει έλλειψη διαθεσιμότητας για τη λήψη ενός συνόλου δεδομένων αναφοράς. Η απουσία βασικής αλήθειας επιδεινώνει τη δυσκολία να τεθεί το πρόβλημα ως μια εποπτευόμενη εργασία μηχανικής μάθησης.
- **Ανάλυση περιεχομένου:** Μία από τις μεγαλύτερες προκλήσεις είναι η ανάλυση του τεράστιου όγκου πληροφοριών περιεχομένου στα μέσα κοινωνικής δικτύωσης. Για τον εντοπισμό οργανωμένων δραστηριοτήτων crowdurfing, το μεγαλύτερο μέγεθος δειγμάτων μπορεί να οδηγήσει σε καλύτερα και πιο εξακριβωμένα αποτελέσματα.

Ωστόσο, πχ στο Twitter, για παράδειγμα, ο αριθμός των μηνιαίων ενεργών χρηστών ξεπερνά τα 300 εκατομμύρια και ο αριθμός των tweet την ημέρα είναι πάνω από 500 εκατομμύρια.

- **Εξελισσόμενες στρατηγικές:** Οι προσεγγίσεις μηχανικής μάθησης επικεντρώνονται στον εντοπισμό προτύπων δυσμενών δραστηριοτήτων, που μπορεί να είναι δύσκολο να αντιμετωπιστούν με νέες στρατηγικές αναδυόμενων καμπανιών crowdurfing. Για παράδειγμα, οι παραδοσιακές μέθοδοι ανίχνευσης του crowdurfing μπορεί να εκμεταλλεύονται μόνο το περιεχόμενο, τις εικόνες και τα βίντεο που χρησιμοποιούνται όλο και περισσότερο. Επιπλέον, με τις υπάρχουσες μεθόδους θεωρείται συνήθως ότι οι αναρτήσεις των εργαζομένων στο crowdurfing είναι σχεδόν διπλές, καθώς το περιεχόμενο είναι συνήθως προγραμμαμένο από τους υποστηρικτές. Ωστόσο, έχουν βρεθεί επίσης εργαζόμενοι στο crowdurfing, οι οποίοι θα μπορούσαν να συντάξουν πιο πρωτότυπο περιεχόμενο. Η εξελισσόμενη στρατηγική crowdurfing μπορεί εύκολα να παρακάμψει το στατικό φιλτράρισμα των παραδοσιακών προσεγγίσεων ανίχνευσης.

- **Οι εξελισσόμενοι συμμετέχοντες:** Το Crowdsourcing είναι δημοφιλές καθώς παρέχει έναν τρόπο στους ανθρώπους να μετατρέψουν τον ελεύθερο χρόνο τους σε ανταμοιβές δημιουργίας εσόδων. Πολλοί εργαζόμενοι στο crowdurfing εργάζονται ως «μερικής απασχόλησης», επομένως οι αντίστοιχοι λογαριασμοί μέσω κοινωνικής δικτύωσης συνήθως δεν είναι αφιερωμένοι για crowdurfing και περιέχουν πληροφορίες περιεχομένου για τον εαυτό τους αντί για καμπάνιες. Οι προσωπικές αναρτήσεις μπορεί να χρησιμεύσουν ως καμουφλάζ που συγκαλύπτει τις πληροφορίες της εκστρατείας. Επομένως, είναι δύσκολο για τις παραδοσιακές προσεγγίσεις να ανιχνεύουν τέτοιους εξελισσόμενους συμμετέχοντες στο crowdurfing με μικτό περιεχόμενο και θα απαιτούσε λεπτομερή ανάλυση στις αναρτήσεις των χρηστών.

- **Ετερογενή δεδομένα:** Οι πληροφορίες πολυμέσων είναι ολοένα και πιο διαδεδομένες στους ιστότοπους κοινωνικής δικτύωσης. Οι πληροφορίες πολυμέσων παρέχονται από νεοϊδρυθέντες ιστότοπους κοινωνικών μέσων που είναι αφιερωμένοι σε εικόνες και βίντεο όπως το Instagram και το Pinterest. Οι υπάρχοντες ιστότοποι κοινωνικής δικτύωσης όπως το Facebook και το Twitter χρησιμοποιούνται επίσης ευρέως για τη διάδοση εικόνων και σύντομων βίντεο. Επίσης, οι υπάρχοντες ιστότοποι πολυμέσων όπως το Youtube, εξελίσσονται με την προσθήκη περισσότερων κοινωνικών

λειτουργιών. Επομένως, εικόνες, βίντεο κλιπ και κείμενα χρησιμοποιούνται ταυτόχρονα για καμπάνιες. Η ετερογένεια των δεδομένων απαιτεί καλύτερη προσαρμοστικότητα, κάτι που λείπει από τις υπάρχουσες μεθόδους.

Οι διάφορες μελέτες που διεξήχθησαν για την ανίχνευση του crowdurfing έχουν λάβει υπόψη μία ή περισσότερες από αυτές τις προκλήσεις. Παρακάτω, θα παρουσιάσουμε και θα συζητήσουμε τα κύρια ερευνητικά τους ευρήματα.

Πριν εισαγάγουμε τον εντοπισμό δραστηριοτήτων crowdurfing στα μέσα κοινωνικής δικτύωσης, εισάγουμε πρώτα τις ακόλουθες θεμελιώδεις αρχές που έχουν χρησιμοποιηθεί για τη διερεύνηση του crowdurfing:

- Εξόρυξη και δημιουργία προφίλ χρηστών μέσω κοινωνικής δικτύωσης
- Μοντελοποίηση της διάδοσης πληροφοριών στα μέσα κοινωνικής δικτύωσης

4.3 Εξόρυξη και δημιουργία προφίλ χρηστών μέσω κοινωνικής δικτύωσης

Στις πλατφόρμες μέσω κοινωνικής δικτύωσης, οι χρήστες δημοσιεύουν περιεχόμενο, προωθούν αναρτήσεις φίλων και λαμβάνουν πληροφορίες από τους φίλους τους. Οι άνθρωποι αλληλεπιδρούν μεταξύ τους κάνοντας φίλους και ανταλλάσσοντας πληροφορίες. Αυτές οι δραστηριότητες προκαλούν ένα δίκτυο μεταξύ των χρηστών των μέσων κοινωνικής δικτύωσης που έχει μελετηθεί καλά για την κατανόηση των προτύπων του crowdurfing και άλλων δυσμενών επιθέσεων, τα οποία επεξηγούνται παρακάτω (Thai et al., 2016).

- Κοινωνικά δίκτυα: Οι χρήστες των μέσων κοινωνικής δικτύωσης θα μπορούσαν να ακολουθούν ο ένας τον άλλον για να εγγραφούν στις κοινές πληροφορίες άλλων ατόμων. Η συμπεριφορά της παρακολούθησης δημιουργεί συνδέσμους, οι οποίοι καταλήγουν σε μια δομή κοινωνικού δικτύου. Η δομή του κοινωνικού δικτύου έχει χρησιμοποιηθεί για τον εντοπισμό δυσμενών συμπεριφορών στα κοινωνικά δίκτυα. Οι πληροφορίες που έχουν χρησιμοποιηθεί περιλαμβάνουν τον αριθμό των συνδέσμων,

τα περιεχόμενα και τα προφίλ των φίλων και τις δομές και τις ιεραρχίες της κοινότητας που προκαλούνται από το κοινωνικό δίκτυο.

- **Συνομιλίες:** Εκτός από την κοινή χρήση πληροφοριών, οι χρήστες των κοινωνικών δικτύων μπορούν επίσης να σχολιάζουν ο ένας τον άλλον για να μοιράζονται και να ανταλλάσσουν απόψεις, οι οποίες καταλήγουν σε συνομιλίες μεταξύ των χρηστών. Οι ιδέες που εκφράζονται στα σχόλια μπορούν να θεωρηθούν ως δείκτης της ποιότητας των πληροφοριών. Για παράδειγμα, εάν κάποια αναδυόμενη ιστορία είναι απίθανο να είναι αληθινή, οι άνθρωποι συνήθως προσπαθούν να την επαληθεύσουν και να τη διορθώσουν κάνοντας ερωτήσεις ή σχολιάζοντας αρνητικά.
- **Ροές κειμένων:** Στα κοινωνικά δίκτυα, δημιουργείται συνεχώς περιεχόμενο που περιλαμβάνει μια ροή κειμένων. Δεδομένου ότι οι καμπάνιες είναι ευαίσθητες στον χρόνο και μπορεί να εμφανιστούν κατά τη διάρκεια συγκεκριμένων χρονικών περιόδων, ενδέχεται να υπάρχουν συγκεκριμένα μοτίβα, όπως η αδρανοποίηση (χαμηλός όγκος επισκεψιμότητας/δραστηριότητας σε μεγάλο χρονικό διάστημα) και οι απότομες εκρήξεις (υψηλός όγκος επισκεψιμότητας/δραστηριότητας σε σύντομο χρονικό διάστημα). Αυτά τα χρονικά και ποσοτικά πρότυπα μπορούν να είναι ένα καλό μέτρο για την ένδειξη μη φυσιολογικών δραστηριοτήτων.
- **Χρονικά και χωρικά μοτίβα:** Οι ερευνητές πραγματοποίησαν αναλύσεις στις συμπεριφορές των χρηστών για να εντοπίσουν μη φυσιολογικές δραστηριότητες. Ένα θέμα που είναι δημοφιλές στα μέσα κοινωνικής δικτύωσης σχετίζεται συνήθως με συγκεκριμένη ώρα και τοποθεσίες, επομένως η ώρα και η τοποθεσία των χρηστών, σε συνδυασμό με το περιεχόμενο που δημοσίευσαν, μπορούν να ληφθούν ως δείκτης του κανόνα των πληροφοριών.
- **Πρωτοτυπία πληροφοριών:** Η πρωτοτυπία των πληροφοριών περιγράφει τη μοναδικότητα και την πρωτοτυπία των πληροφοριών. Δεδομένου ότι το crowdurfing είναι συνήθως οργανωμένο, το αντίστοιχο περιεχόμενο θα πρέπει να είναι λιγότερο ποικίλο και να έχει πιο περιορισμένο λεξιλόγιο. Αν και η πρωτοτυπία έχει χρησιμοποιηθεί ευρέως για παραδοσιακές προσεγγίσεις για τον εντοπισμό του astroturf, στο πλαίσιο της ανίχνευσης του crowdurfing, ωστόσο, η μέτρηση της πρωτοτυπίας των πληροφοριών είναι πιο δύσκολη καθώς ο όγκος των πληροφοριών

είναι τεράστιος και πολλοί άνθρωποι μπορεί να δημοσιεύουν περιεχόμενο άλλων ατόμων χωρίς αναφορά.

4.4 Μοντελοποίηση διάχυσης πληροφοριών στα μέσα κοινωνικής δικτύωσης

Η δημοσίευση νέου περιεχομένου και η διάδοση, η προώθηση πληροφοριών από άλλα άτομα σε φίλους και η σύνταξη σχολίων, αυτές οι δραστηριότητες είναι επιτακτικές για τους χρήστες των μέσων κοινωνικής δικτύωσης. Οι δραστηριότητες που απεικονίζουν αυτούς τους ανθρώπους ως ηθοποιούς στα κοινωνικά δίκτυα και τους βοηθούν να επηρεάσουν άλλα άτομα έχουν ως αποτέλεσμα τη διάχυση πληροφοριών μεταξύ των χρηστών των μέσων κοινωνικής δικτύωσης με την πάροδο του χρόνου. Παρακάτω σημειώνονται οι θεμελιώδεις αρχές της διάχυσης πληροφοριών και το πώς τα μοντέλα διάχυσης πληροφοριών μπορούν να προσαρμοστούν για να αντιμετωπίσουν τις πληροφορίες του crowdurfing.

- Μοντέλα διάχυσης γενικών πληροφοριών: Τα βασικά μοντέλα διάχυσης πληροφοριών περιλαμβάνουν το μοντέλο SIR, το μοντέλο σημείου ανατροπής, το ανεξάρτητο μοντέλο αλληλουχίας και το γραμμικό μοντέλο. Τα μοντέλα διάχυσης πληροφοριών εξηγούν τις πληροφορίες μελετώντας την τοπολογία του δικτύου και τον μηχανισμό ανταλλαγής πληροφοριών. Το ανεξάρτητο μοντέλο αλληλουχίας έχει υιοθετηθεί για πληροφορίες μέσων κοινωνικής δικτύωσης, καθώς λαμβάνει υπόψη τις σχέσεις και την επιρροή των φίλων. Το ανεξάρτητο μοντέλο αλληλουχίας έχει επίσης χρησιμοποιηθεί για τη μοντελοποίηση της διάχυσης πληροφοριών μεταξύ των χρηστών κοινωνικών δικτύων υποθέτοντας ότι οι σχέσεις μεταξύ των ανθρώπων είναι ανεξάρτητες (Lim et al., 2011).

- Διάχυση πληροφοριών για το crowdurfing: Προκειμένου να προσαρμοστούν τα γενικά μοντέλα διάχυσης πληροφοριών στη διάχυση πληροφοριών από το crowdurfing, μπορούν να εισαχθούν διαφορετικά είδη ανταλλαγής πληροφοριών και περισσότεροι ρόλοι. Για παράδειγμα, το ανεξάρτητο μοντέλο αλληλουχίας υποθέτει ότι η ανταλλαγή πληροφοριών οδηγεί στην ενημέρωση, ωστόσο, για την

παραπληροφόρηση από το crowdfunding, η «ανταλλαγή» δεν αφορά μόνο την ενημέρωση, αλλά περισσότερο το να πείσει άλλους ανθρώπους να εμπιστευτούν τις πληροφορίες. Ως εκ τούτου, οι ερευνητές εισάγουν επίπεδα πεποιθήσεων στη διαδικασία ανταλλαγής πληροφοριών για να μελετήσουν πώς διαδίδονται οι πληροφορίες. Ένας άλλος τρόπος είναι να εισαχθούν περισσότεροι ρόλοι (Tian et al., 2015).

4.5 Προσεγγίσεις που βασίζονται στο περιεχόμενο

Προκειμένου να εντοπιστεί το crowdfunding, ένας πιθανός τρόπος είναι να εξεταστεί το περιεχόμενο των μέσων κοινωνικής δικτύωσης και να εντοπιστούν παρόμοιες και διπλές αναρτήσεις. Για παράδειγμα, οι Wang et al. (2013) δημιούργησαν μοντέλα συμπεριφοράς εκμεταλλευόμενοι τα περιεχόμενα του χρήστη. Ομοίως, οι Yang et al. (2015) επιχείρησαν να ανακαλύψουν τα θέματα στα οποία επικεντρώνονται οι εργαζόμενοι στο crowdfunding. Όπως αναφέρθηκε προηγουμένως, οι πελάτες του crowdfunding μπορεί να ζητήσουν από τους εργαζόμενους στο crowdfunding να συντάξουν πρωτότυπο περιεχόμενο, επομένως οι αντίστοιχες αναρτήσεις ενδέχεται να μην μοιράζονται παρόμοια μοτίβα που οδηγούν σε διακριτές «συμπεριφορικές υπογραφές». Οι αναρτήσεις που φαίνεται να είναι διαφορετικές μοιράζονται τον ίδιο στόχο να οδηγήσουν άλλους χρήστες σε έναν συγκεκριμένο πόρο, όπως έναν λογαριασμό κάποιου χρήστη ή μια εξωτερική ιστοσελίδα. Επομένως, οι «στόχοι» των αναρτήσεων μπορούν να χρησιμοποιηθούν ως δείκτης συστηματικά οργανωμένης παραγωγής περιεχομένου.

4.6 Προσεγγίσεις που βασίζονται στη συμπεριφορά

Δεδομένου ότι οι εργασίες crowdfunding συνήθως απαιτείται να γίνονται σύμφωνα με κανόνες που έχουν καταρτιστεί από τους πελάτες, οι συμπεριφορές των χρηστών των μέσων κοινωνικής δικτύωσης που εμπλέκονται με το crowdfunding είναι χρήσιμες για την αποκάλυψη των μη φυσιολογικών δραστηριοτήτων. Για παράδειγμα, το χρονικό

Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ

μοτίβο μεταφόρτωσης και κλικ στις διευθύνσεις URL μπορεί να χρησιμοποιηθεί για τη διαφοροποίηση του crowdurfing (Wu & Liu, 2018).

4.7 Προσεγγίσεις που βασίζονται στη διάχυση

Οι προσεγγίσεις που βασίζονται στη διάχυση στοχεύουν στην τροποποίηση των υπαρχόντων μοντέλων διάχυσης πληροφοριών για την κατανόηση και τον εντοπισμό δυσμενών δραστηριοτήτων. Μέσω της επέκτασης του ανεξάρτητου μοντέλου αλληλουχίας και του γραμμικού μοντέλου, τα μοντέλα διάχυσης πληροφοριών μπορούν να χρησιμοποιηθούν για να αναπαραστήσουν τη διαδικασία δυσμενών επιθέσεων σε ιστότοπους μέσω κοινωνικής δικτύωσης. Τα μοντέλα διάχυσης μπορούν όχι μόνο να διευκολύνουν την ανίχνευση του crowdurfing αλλά και να βοηθήσουν στην παρέμβαση της διαδικασίας και στην ελαχιστοποίηση των επακόλουθων επιπτώσεων της παραπληροφόρησης. Οι Ratkiewicz et al. (2011) οπτικοποίησαν τη διαδικασία της διάχυσης πληροφοριών, η οποία βοηθά στην κατανόηση του μηχανισμού διάχυσης της καμπάνιας crowdurfing και μπορεί να παρέχει μια φιλική προς τον χρήστη διεπαφή για τη μη αυτόματη επισήμανση ροών περιεχομένου crowdurfing.

4.8 Σύνολα δεδομένων

Ερευνητές και επαγγελματίες έχουν χρησιμοποιήσει διαφορετικά σύνολα δεδομένων για την ανίχνευση του crowdurfing. Ένα σύνολο δεδομένων astroturfing είναι διαθέσιμο για ροές κειμένου στο Twitter, το οποίο είναι δημόσια διαθέσιμο. Το σύνολο δεδομένων αποτελείται από μίμια που λαμβάνονται από το Twitter και περιέχει δεδομένα βασικής αλήθειας τα οποία επισημαίνονται χειροκίνητα. Ένα σύνολο δεδομένων κακόβουλου χρήστη είναι διαθέσιμο για τη μελέτη συμπεριφορών πιθανών εργαζομένων στο crowdurfing. Η βασική αλήθεια είναι διαθέσιμη με το σύνολο δεδομένων, το οποίο επισημαίνεται αυτόματα χρησιμοποιώντας social honeypots. Προκειμένου να μελετηθούν οι συμπεριφορές των χρηστών από διαφορετικά κοινωνικά δίκτυα, περισσότερα σύνολα δεδομένων είναι δημόσια διαθέσιμα στο χώρο

αποθήκευσης δεδομένων κοινωνικών μέσων της ASU. Είναι επίσης δυνατή η απόκτηση δεδομένων crowdurfing μέσω των ιστοσελίδων crowdurfing (πχ Zhubajie, Sandaha και Fiverr) (Zafarani et al., 2014).

Βασικές εφαρμογές (Wu & Liu, 2018):

- Το TweetTracker είναι ένα εργαλείο ανάλυσης που έχει σχεδιαστεί για την παρακολούθηση, ανάλυση και παρακολούθηση πληροφοριών περιεχομένου στα μέσα κοινωνικής δικτύωσης. Το εργαλείο ενσωματώνει τη δυνατότητα τοποθεσίας και μπορεί να βοηθήσει ερευνητές και επαγγελματίες με τη συλλογή δεδομένων βάσει συμβάντων σε δεδομένα πολυμέσων. Το TweetTracker έχει μια ενσωματωμένη μονάδα ανίχνευσης κακόβουλων χρηστών που βασίζονται σε συμβάντα, η οποία μπορεί να χρησιμοποιηθεί για την αναγνώριση συμμετεχόντων στην προπαγάνδα. Η λειτουργία παρακολούθησης μπορεί να χρησιμοποιηθεί για την ανίχνευση του crowdurfing. Το εργαλείο είναι δωρεάν για χρήση για ακαδημαϊκούς σκοπούς κατόπιν έγκρισης.
- Το Truthy είναι ένα εργαλείο που έχει ως βάση του το Twitter, το οποίο στοχεύει να ανιχνεύσει την εξάπλωση των astroturfs στα μέσα κοινωνικής δικτύωσης εξετάζοντας κακόβουλους λογαριασμούς. Αξίζει να σημειωθεί πως είναι δωρεάν το να ελέγξει κανείς έναν ατομικό λογαριασμό από το εργαλείο.
- Το Debot είναι ένα δωρεάν διαθέσιμο εργαλείο για τον έλεγχο των χρηστών bot στα μέσα κοινωνικής δικτύωσης. Δεδομένου ότι το εργαλείο εστιάζει σε συσχετισμένες και σχεδόν διπλότυπες δραστηριότητες διαφορετικών χρηστών μέσων κοινωνικής δικτύωσης, θα μπορούσε να χρησιμοποιηθεί για τον εντοπισμό εργαζομένων στο crowdurfing που είναι οργανωμένοι για μια καμπάνια. Οι κωδικοί και τα δεδομένα είναι ελεύθερα διαθέσιμα.

4.8 Μελλοντικές κατευθύνσεις

Παραπάνω αναφέρθηκαν σημαντικές προκλήσεις, θεμελιώδεις αρχές και βασικές έννοιες της ανίχνευσης του crowdurfing. Εστίασαμε σε αρκετές αξιολογικές εφαρμογές και σύνολα δεδομένων που προσφέρουν αξιοσημείωτες ευκαιρίες για περαιτέρω έρευνα. Αν και το θέμα έχει μελετηθεί εκτενώς, οραματιζόμαστε διάφορες κατευθύνσεις για μελλοντική εργασία, καθώς η εξάρτησή μας από τα μέσα κοινωνικής δικτύωσης όλο και μεγαλώνει. Με περισσότερες ιστοσελίδες μέσω κοινωνικής

δικτύωσης να δημιουργούνται και να είναι δημοφιλείς, μια εκστρατεία crowdurfing θα μπορούσε να ξεκινήσει ταυτόχρονα σε διαφορετικούς ιστότοπους κοινωνικών μέσων. Επίσης, με τον αυξανόμενο ρόλο των μέσων κοινωνικής δικτύωσης σε γεγονότα του πραγματικού κόσμου, η περαιτέρω έρευνα θα συνεπαγόταν όχι μόνο διαφορετικά κανάλια μέσων κοινωνικής δικτύωσης αλλά και τα αποτελέσματα γεγονότων που συμβαίνουν στον πραγματικό κόσμο. Για παράδειγμα, μια εταιρεία μπορεί να ξεκινήσει εκστρατείες κατά τη διάρκεια μιας κρίσης δημοσίων σχέσεων. Η ενσωμάτωση της εξέτασης αυτών των πραγματικών γεγονότων μπορεί να διευκολύνει περαιτέρω τον εντοπισμό του crowdurfing (Wu & Liu, 2018).

ΚΕΦΑΛΑΙΟ 5^ο

5.1 Επιθέσεις crowdsourced σε συστήματα αξιολόγησης e-commerce

Οι περισσότεροι δημοφιλείς ιστότοποι ηλεκτρονικού εμπορίου σήμερα βασίζονται στα σχόλια των χρηστών για την αξιολόγηση προϊόντων, υπηρεσιών και διαδικτυακού περιεχομένου. Τα σχόλια του crowdsourced περιλαμβάνουν συνήθως μια κριτική ή γνώμη, που περιγράφει την εμπειρία ενός χρήστη από ένα προϊόν ή μια υπηρεσία, μαζί με μια βαθμολογία αξιολόγησης (συνήθως σε κλίμακα από το 1 έως το 5). Δυστυχώς, πολλά συστήματα κριτικής μαστίζονται από ψεύτικες κριτικές, π.χ. Yelp, Amazon, iTunes και TripAdvisor, όπου ένας εισβολέας χειραγωγεί τη γνώμη του πλήθους χρησιμοποιώντας ψεύτικες ή παραπλανητικές κριτικές. Για να ενισχύσουν τη φήμη τους ή να βλάψουν τη φήμη ενός ανταγωνιστή, οι επιχειρήσεις μπορούν να ζητήσουν ψεύτικες κριτικές που εκφράζουν μια υπερβολικά θετική ή αρνητική γνώμη για μια επιχείρηση (Lee et al., 2014).

Μελέτες για το Yelp διαπίστωσαν ότι μια αύξηση κατά ένα αστέρι για τα εστιατόρια μπορεί να οδηγήσει σε 5-9% αύξηση των εσόδων. Ιστότοποι όπως το Yelp και το Amazon έχουν εμπλακεί με συνέπεια σε μια μάχη «γάτας και ποντικιού» με ψεύτικες κριτικές, καθώς οι επιτιθέμενοι προσπαθούν να προσαρμόσουν και να παρακάμψουν διάφορα αμυντικά συστήματα. Το σύστημα φίλτρου αναθεώρησης του Yelp επισημαίνει ύποπτες κριτικές και προειδοποιεί ακόμη και τον καταναλωτή εάν μια επιχείρηση είναι ύποπτη ότι εμπλέκεται σε μεγάλης κλίμακας χειραγώγηση γνώμης. Πρόσφατα, οι επιθέσεις είναι γνωστό ότι δημιουργούν εξαιρετικά παραπλανητικές ψεύτικες κριτικές που γράφτηκαν από πληρωμένους χρήστες. Πολλά από αυτά προέρχονται από κακόβουλες αγορές crowdsourcing, γνωστές ως συστήματα crowdurfing, όπου μια μεγάλη ομάδα ανθρώπων εργαζομένων παρέχει προσπάθεια κατ' απαίτηση για την ολοκλήρωση διαφόρων κακόβουλων εργασιών (Roberts, 2015).

5.2 Αξιολόγηση ποιότητας κριτικών που έχουν δημιουργηθεί από μηχανήματα

Σε αυτήν την ενότητα, αξιολογούμε την ποιότητα των κριτικών που δημιουργούνται από μηχανή σε δύο διαστάσεις. Αρχικά, διερευνάται εάν οι δημιουργούμενες αξιολογήσεις μπορούν να παρακάμψουν τον εντοπισμό από υπάρχουσες αλγοριθμικές προσεγγίσεις. Δεύτερον, παρουσιάζονται κριτικές εστιατορίων που είναι και ψεύτικες και πραγματικές. Στόχος μας είναι να κατανοήσουμε εάν οι άνθρωποι μπορούν να διακρίνουν τις ψεύτικες από τις πραγματικές κριτικές.

Οι Yao et al. (2014) εστίασαν σε δύο δημοφιλείς αλγοριθμικές τεχνικές για τη διάκριση των κριτικών που δημιουργούνται από μηχανή από τις πραγματικές κριτικές: (1) ένα εποπτευόμενο σχήμα ML που βασίζεται σε γλωσσικά χαρακτηριστικά, (2) έναν ανιχνευτή λογοκλοπής για τον έλεγχο διπλών κριτικών.

Φίλτρο αναθεώρησης που βασίζεται σε ML

Η χρήση ταξινομητών μηχανικής μάθησης για τον εντοπισμό ψεύτικων κριτικών είναι ένα καλά μελετημένο πρόβλημα. Οι περισσότερες από τις προηγούμενες μελέτες βασίζονται στην παρατήρηση ότι τα χαρακτηριστικά των ψεύτικων κριτικών αποκλίνουν από τις πραγματικές κριτικές σε πολλές γλωσσικές διαστάσεις. Από τους Yao et al. (2014) προσδιορίστηκαν 5 ομάδες γλωσσικών χαρακτηριστικών, αποτελούμενες από 77 συνολικά χαρακτηριστικά που προηγουμένως επέδειξαν ισχυρή διακριτική δύναμη για τη διάκριση ψεύτικων και πραγματικών κριτικών:

- Χαρακτηριστικό ομοιότητας (1): Καταγράφει ομοιότητα μεταξύ των προτάσεων σε μια κριτική σε επίπεδο λέξης (Rayana & Akoglu, 2015).
- Δομικά χαρακτηριστικά (4): Αποτυπώνουν τις δομικές πτυχές μιας ανασκόπησης. Τα μεμονωμένα χαρακτηριστικά περιλαμβάνουν τον αριθμό των λέξεων, τον αριθμό των προτάσεων, το μέσο μήκος πρότασης (λέξεις) και το μέσο μήκος λέξης (χαρακτήρων) (Rayana & Akoglu, 2015).
- Συντακτικά χαρακτηριστικά (6): Καταγράφουν τις γλωσσικές ιδιότητες της κριτικής με βάση την επισήμανση τμημάτων του λόγου (parts-of-speech - POS). Τα

χαρακτηριστικά περιλαμβάνουν (ξεχωριστά) ποσοστά ουσιαστικών, ρημάτων, επιθέτων και επιρρημάτων, προσωπικές ανωνυμίες (Rayana & Akoglu, 2015).

- Σημασιολογικά χαρακτηριστικά (4): Αποτυπώνουν την υποκειμενικότητα και το συναίσθημα των κριτικών. Τα χαρακτηριστικά περιλαμβάνουν ποσοστό υποκειμενικών λέξεων, ποσοστό αντικειμενικών λέξεων, ποσοστό θετικών λέξεων και ποσοστό αρνητικών λέξεων. Όλα αυτά τα χαρακτηριστικά ορίζονται στο SentiWordNet, μια δημοφιλή λεξιλογική πηγή για την εξόρυξη γνώμης (Rayana & Akoglu, 2015).

- Χαρακτηριστικά LIWC (62): Το λογισμικό Linguistic Inquiry and Word Count (LIWC) είναι ένα ευρέως χρησιμοποιούμενο εργαλείο ανάλυσης κειμένου στις κοινωνικές επιστήμες. Κατηγοριοποιεί ~4.500 λέξεις-κλειδιά σε ~68 ψυχολογικές τάξεις (π.χ. γλωσσικές διαδικασίες, ψυχολογικές διεργασίες, προσωπικές ανησυχίες και προφορικές κατηγορίες). Χρησιμοποιείται το ποσοστό του πλήθους λέξεων σε κάθε ομάδα ως χαρακτηριστικό και αποκλείονται τα χαρακτηριστικά που περιλαμβάνονται ήδη στις προηγούμενες ομάδες (Rayana & Akoglu, 2015).

Οι Yao et al. (2014) εκπαίδευσαν έναν γραμμικό ταξινομητή SVM στο σύνολο δεδομένων Yelp ground-truth, που αποτελείται από πραγματικές κριτικές (μη φιλτραρισμένες κριτικές Yelp) και ψεύτικες κριτικές (φιλτραρισμένες κριτικές Yelp). Μετά την εκπαίδευση και με τα 77 γλωσσικά χαρακτηριστικά, δοκίμασαν την απόδοση του ταξινομητή στο σύνολο δεδομένων επίθεσης Yelp, που αποτελείται από πραγματικές κριτικές και κριτικές που δημιουργούνται από μηχανή και εκτέλεσαν 10πλάσια διασταυρούμενη επικύρωση, αναφέροντας τη μέση απόδοση. Η αξιολόγηση της απόδοσης επίθεσης χρησιμοποιεί ακρίβεια και ανάκληση.

Ανιχνευτής λογοκλοπής

Η επίτευξη λογικής γλωσσικής ποιότητας δεν αποκλείει το ενδεχόμενο να είναι ψεύτικη η κριτική. Μια απλή επίθεση περιλαμβάνει τη δημιουργία ψεύτικων κριτικών με αντιγραφή ή μερική αντιγραφή από πραγματικές κριτικές. Σε τέτοιες περιπτώσεις, η ποιότητα της κριτικής θα ήταν αρκετά καλή και θα περνούσε το γλωσσικό φίλτρο. Η τυπική λύση είναι να βασιστούμε σε ελεγκτές λογοκλοπής για τον εντοπισμό των διπλών ή σχεδόν διπλότυπων κριτικών. Δεδομένου ότι το μοντέλο RNN έχει

εκπαιδευτεί να δημιουργεί κείμενο παρόμοιο, εξετάζεται εάν οι αξιολογήσεις που δημιουργούνται από μηχανή είναι διπλότυπες ή σχεδόν διπλότυπες κριτικών. Για τη διεξαγωγή ελέγχου λογοκλοπής, υποθέτεται ότι ο πάροχος υπηρεσιών έχει πρόσβαση σε μια βάση δεδομένων με κριτικές που χρησιμοποιείται για την «εκπαίδευση» του RNN. Στη συνέχεια, λαμβάνοντας υπόψη μια αναθεώρηση που δημιουργείται από μηχανή, ο πάροχος υπηρεσιών εκτελεί έναν έλεγχο λογοκλοπής συγκρίνοντάς τον με κριτικές στη βάση δεδομένων. Αυτό είναι το καλύτερο σενάριο για ένα τεστ λογοκλοπής και μας βοηθά να κατανοήσουμε τις δυνατότητές του να ανιχνεύει κριτικές που δημιουργούνται (Yao et al., 2014).

Το Winnowing αποτελεί μια ευρέως χρησιμοποιούμενη μέθοδο για τον εντοπισμό διπλού ή σχεδόν διπλού κειμένου. Για ένα ύποπτο κείμενο, το Winnowing δημιουργεί πρώτα ένα σύνολο «αποτυπωμάτων» και στη συνέχεια συγκρίνει τα «αποτυπώματα» μεταξύ του ύποπτου κειμένου και του κειμένου στη βάση δεδομένων. Η ομοιότητα μεταξύ δύο αξιολογήσεων υπολογίζεται χρησιμοποιώντας το Jaccard Similarity των αποτυπωμάτων τους που δημιουργήθηκαν από το Winnowing. Η βαθμολογία ομοιότητας λογοκλοπής για μια μεμονωμένη αξιολόγηση υπολογίζεται ως η μέγιστη ομοιότητα με όλες τις άλλες αξιολογήσεις στο σύνολο δεδομένων και κυμαίνεται από 0 έως 1 (το 1 υποδηλώνει πανομοιότυπες κριτικές) (Schleimer et al., 2003).

Στην έρευνα των Yao et al. (2014) επιλέχθηκε ένα τυχαίο δείγμα 10.000 κριτικών που δημιουργούνται από μηχανή για το τεστ λογοκλοπής και η βάση δεδομένων περιλαμβάνει ολόκληρο το εκπαιδευτικό σύνολο δεδομένων Yelp. Επιπλέον, οι κριτικές που δημιουργούνται από μηχανή και οι πραγματικές κριτικές εμφανίζουν παρόμοιες βαθμολογίες λογοκλοπής, καθιστώντας έτσι πιο δύσκολη τη διάκρισή τους. Επομένως, παραμένει δύσκολο να ανιχνευθούν κριτικές που δημιουργούνται από μηχανή χρησιμοποιώντας έναν έλεγχο λογοκλοπής χωρίς να επισημανθεί κατά λάθος μεγάλο αριθμό πραγματικών κριτικών. Αυτό δείχνει ότι το RNN δεν αντιγράφει απλώς τις υπάρχουσες κριτικές (Yao et al., 2014).

Ανεξάρτητα από το πόσο καλά αποδίδουν οι αξιολογήσεις που δημιουργούνται από μηχανή σε στατιστικά μέτρα και δοκιμές, το πραγματικό τεστ είναι εάν μπορούν να θεωρηθούν ως πραγματικές κριτικές όταν διαβάζονται από ανθρώπους. Στην πράξη, οι πάροχοι υπηρεσιών είναι γνωστό ότι εμπλέκουν ανθρώπινους συντονιστές

περιεχομένου για να διαχωρίσουν τις κριτικές που δημιουργούνται από μηχανή από τις πραγματικές κριτικές (Slee, 2014).

Για να μετρηθεί η ανθρώπινη απόδοση, οι Yao et al. (2014) διεξήγαγαν τις έρευνές τους στο Amazon Mechanical Turk (AMT). Κάθε έρευνα περιλαμβάνει ένα όνομα εστιατορίου, περιγραφή (εξηγώντας την κατηγορία του εστιατορίου και την περιγραφή που παρέχεται από την επιχείρηση στο Yelp) και ένα σύνολο κριτικών, το οποίο περιλαμβάνει κριτικές που δημιουργούνται από μηχανή και πραγματικές κριτικές που γράφτηκαν για αυτό το εστιατόριο. Στη συνέχεια, ζήτησαν από κάθε εργαζόμενο να επισημάνει τις κριτικές που θεωρεί ότι είναι ψεύτικες, χρησιμοποιώντας οποιαδήποτε βάση για την κρίση του. Για την έρευνα, επιλέχθηκαν 40 εστιατόρια με τον μεγαλύτερο αριθμό κριτικών στο σύνολο δεδομένων βασικής αλήθειας. Για κάθε εστιατόριο, διεξήγαγαν έρευνες, καθεμία από τις οποίες περιλαμβάνει 20 τυχαίες κριτικές, εκ των οποίων ένα μέρος (X) είναι κριτικές που δημιουργούνται από μηχανή και οι υπόλοιπες είναι πραγματικές κριτικές από το Yelp. Ο αριθμός X επιλέχθηκε τυχαία μεταξύ 0 και 5, έτσι ώστε η αναμενόμενη αναλογία ψεύτικων κριτικών (12,5%) να ταιριάζει με τη ρύθμιση του πραγματικού κόσμου (11,99%). Επιπλέον, ελέγχθηκε η ποιότητα των πραγματικών κριτικών που εμφανίζονται στις έρευνες για να καλύψουμε όλο το φάσμα της χρησιμότητας. Αξιοποιήθηκε η χρησιμότητα της κριτικής (μια απλή καταμέτρηση του αριθμού των χρηστών που βρήκαν την κριτική χρήσιμη) τα μεταδεδομένα που παρέχονται από το Yelp για κάθε κριτική.

Σύμφωνα με τα αποτελέσματα της μελέτης των Yao et al. (2014) η μέση βαθμολογία χρησιμότητας των ψευδών αρνητικών (χωρίς επισήμανση κριτικών που δημιουργούνται από μηχανή) είναι κοντά σε αυτήν των αληθινών αρνητικών (χωρίς επισήμανση πραγματικών κριτικών Yelp). Οι κριτικές που δημιουργούνται από μηχανή έχουν μέση βαθμολογία χρησιμότητας 3,15, που είναι κοντά στη μέση βαθμολογία χρησιμότητας 3,28 για πραγματικές κριτικές Yelp. Δηλαδή, οι εργαζόμενοι σκέφτονται τις κριτικές που δημιουργούνται από μηχανή ως σχεδόν εξίσου χρήσιμες με τις πραγματικές κριτικές. Συνολικά, κατέστη σαφές πως οι κριτικές που δημιουργούνται από μηχανή καταφέρνουν να μιμούνται την ποιότητα των πραγματικών κριτικών.

5.3 Άμυνα έναντι κριτικών που δημιουργούνται από μηχανή

Σε αυτήν την ενότητα, προτείνουμε ένα εποπτευόμενο σχήμα για τον εντοπισμό αξιολογήσεων που δημιουργούνται από μηχανή με υψηλή ακρίβεια.

Υπόθεση: Υποθέτουμε ότι ο πάροχος υπηρεσιών έχει πρόσβαση σε ένα περιορισμένο σύνολο κριτικών που δημιουργούνται από το μοντέλο γλώσσας ενός εισβολέα και σε ένα σύνολο πραγματικών κριτικών που διατίθενται στον ιστότοπο αξιολόγησης.

Ουσιαστικά, προσπαθούμε να αναπτύξουμε έναν ταξινομητή μηχανικής μάθησης ικανό να ανιχνεύει την έξοδο ενός άλλου μοντέλου ML. Σε ένα ιδανικό σενάριο, αυτό φαίνεται αδύνατο, επειδή το μοντέλο της γεννήτριας και ο ανιχνευτής είναι πιθανό να χρησιμοποιούν τις ίδιες ακριβώς μετρήσεις στις ίδιες εισόδους (ή σε σύνολα δεδομένων εκπαίδευσης). Έτσι, οποιαδήποτε μέτρηση που ένας ανιχνευτής ML είναι ικανός να ανιχνεύσει μπορεί να ληφθεί υπόψη από τη γεννήτρια που βασίζεται σε ML (Song et al., 2015).

Αν και αναμένουμε ότι ένας εισβολέας γνωρίζει οποιαδήποτε μέτρηση που χρησιμοποιείται από τον ανιχνευτή ML για ανίχνευση, μπορούμε να βρούμε μια αρκετά «σύνθετη» μέτρηση όπου η ακριβής καταγραφή (και αναπαραγωγή) της μέτρησης απαιτεί μια εξαιρετικά μεγάλη υποδομή RNN που είναι πέρα από τις δυνατότητες των περισσότερων επιτιθέμενων. Αυτό αξιοποιεί το γεγονός ότι ένα μοντέλο παραγωγής γλώσσας δημιουργεί μια αναπαράσταση σταθερής μνήμης ολόκληρου του εκπαιδευτικού σώματος, η οποία περιορίζει τον όγκο των πληροφοριών που μπορούν να ληφθούν. Πιο συγκεκριμένα, παρατηρούμε ότι το κείμενο που παράγεται φυσικά (π.χ. από έναν άνθρωπο) αποκλίνει από το κείμενο που δημιουργείται από μηχανή όταν συγκρίνουμε την κατανομή επιπέδου χαρακτήρων, ακόμη και όταν τα γλωσσικά χαρακτηριστικά υψηλότερου επιπέδου (π.χ. συντακτικά, σημασιολογικά χαρακτηριστικά) μπορεί να είναι παρόμοια. Αυτή η απόκλιση στην κατανομή επιπέδου χαρακτήρων οφείλεται κυρίως στην απώλεια πληροφοριών που είναι εγγενής στη διαδικασία παραγωγής μηχανών, η οποία είναι συνάρτηση της ισχύος μοντελοποίησης του χρησιμοποιούμενου RNN (Song et al., 2015).

Σημειώνουμε ότι η μέτρηση κατανομής επιπέδου χαρακτήρων είναι κατάλληλη για τους σκοπούς μας, επειδή είναι η μέτρηση χαμηλότερου επιπέδου που μοντελοποιείται από το δικό μας RNN και επομένως είναι πολύ πιθανό να γίνει εμπόδιο πολυπλοκότητας για το RNN. Ένας εισβολέας μπορεί να σκεφτεί μια γεννήτρια RNN που εκπαιδεύεται χρησιμοποιώντας διανομές σε επίπεδο λέξης. Ευτυχώς, οι διανομές σε επίπεδο λέξης είναι πιο περίπλοκες και δύσκολες στη μοντελοποίηση, τόσο επειδή ο αριθμός των λέξεων είναι συνδυαστικά μεγαλύτερος από τον αριθμό των έγκυρων χαρακτήρων στην αγγλική γλώσσα, όσο και επειδή ένα τέτοιο μοντέλο θα πρέπει να προσθέσει πρόσθετους κανόνες για έγκυρα σημεία στίξης. Ως εκ τούτου, μια γεννήτρια RNN που στοχεύει διανομές σε επίπεδο λέξης θα ήταν ακόμη πιο περιορισμένη υπολογιστικά (άρα θα προκαλέσει περισσότερη απώλεια πληροφοριών) και διαισθητικά, η άμυνα θα λειτουργούσε τουλάχιστον εξίσου καλά με ένα RNN σε επίπεδο χαρακτήρων (Song et al., 2015).

Υπάρχει ένας αριθμός τεχνικών παραγωγής φυσικής γλώσσας που βασίζονται σε προκαθορισμένα πρότυπα. Αυτά τα συστήματα απαιτούν συνήθως κάποιες γνώσεις τομέα και καλά σχεδιασμένους κανόνες. Οι προσεγγίσεις που βασίζονται στη μάθηση έγιναν δημοφιλείς, δηλαδή η δημιουργία ενός στατιστικού μοντέλου για την εκμάθηση των γλωσσικών πληροφοριών. Σε αυτές τις περιπτώσεις, η ποιότητα του κειμένου που δημιουργείται συσχετίζεται σε μεγάλο βαθμό με την ποιότητα του μοντέλου. Προηγούμενες εργασίες δείχνουν ότι τα καλά εκπαιδευμένα μοντέλα RNN ξεπερνούν τα μοντέλα απλούστερης γλώσσας όπως το N-gram και τα μοντέλα γλώσσας που βασίζονται σε RNN έχουν εμφανιστεί ως μια πολλά υποσχόμενη προσέγγιση για τη δημιουργία κειμένου. Οι ερευνητές πέτυχαν επίσης επιτυχημένα αποτελέσματα στη δημιουργία κειμένου για διαφορετικούς τομείς, συμπεριλαμβανομένων των απαντήσεων μέσω email, της περιγραφής εικόνας και των διαδικτυακών συνομιλιών μέσω κοινωνικών δικτύων (Shang et al., 2015).

Οι Lipton et al. (2015) μελέτησαν επίσης τη δημιουργία αναθεώρησης προϊόντος χρησιμοποιώντας ένα RNN, χωρίς να λάβουν υπόψη μια αντίθετη ρύθμιση. Ο Hovy (2016) πραγματοποίησε μια προκαταρκτική διερεύνηση της δημιουργίας ανασκόπησης με βάση n-gram σε ένα αντίθετο περιβάλλον. Σε αντίθεση με την έρευνα των Yao et al. (2014), δεν θεωρεί ότι τα πιο εξελιγμένα μοντέλα παραγωγής (RNN) αποφεύγουν την

ανίχνευση και δεν εξετάζουν καμία ισχυρή άμυνα ή αντίμετρα. Τέλος, ο Lappas (2012) ανέλυσε τις ψεύτικες κριτικές από την οπτική γωνία του εισβολέα για να προσδιορίσει τους παράγοντες που επιτρέπουν μια επιτυχημένη επίθεση, αλλά δεν έλαβε υπόψη την αυτόματη δημιουργία κριτικών.

5.4 Νευρωνικά δίκτυα και ασφάλεια

Οι περισσότερες προηγούμενες μελέτες σχετικά με την εφαρμογή των νευρωνικών δικτύων επικεντρώνονται στη βελτίωση των υφιστάμενων αμυντικών συστημάτων έναντι διαφόρων ευπαθειών ασφάλειας δικτύων και ιστού. Οι εργασίες σε αυτήν την κατηγορία περιλαμβάνουν κυρίως προτάσεις για τη βελτίωση της ανίχνευσης εισβολής στο δίκτυο, την ταξινόμηση κακόβουλου λογισμικού και συστήματα ελέγχου ταυτότητας με βάση κωδικούς πρόσβασης. Παρόμοια με την εργασία των Yao et al. (2014), μερικές μελέτες διερεύνησαν επίσης τη σκοπιμότητα επίθεσης σε διαδικτυακά συστήματα χρησιμοποιώντας DNN. Αυτό περιλαμβάνει προτάσεις για αυτόματη επίλυση CAPTCHA χρησιμοποιώντας Συνελκτικά Νευρωνικά Δίκτυα και αυτόματη δημιουργία τομέων κακόβουλου λογισμικού χρησιμοποιώντας Generative Adversarial Networks (Anderson et al., 2016).

Προηγούμενες εργασίες μελέτησαν αγορές crowdurfing που παρέχουν ανθρώπινο εργατικό δυναμικό για να επιτρέψουν επιθέσεις σε μια ποικιλία διαδικτυακών πλατφορμών, όπως συστήματα αναθεώρησης, μέσα κοινωνικής δικτύωσης και μηχανές αναζήτησης. Η μελέτη των Wang et al. (2014) διερεύνησε τον εντοπισμό κακόβουλων εργαζομένων στο πλήθος που χρησιμοποιούν μηχανική μάθηση και διερευνά την ευρωστία των ταξινομητών μηχανικής μάθησης έναντι των τακτικών αποφυγής από τους εργάτες του πλήθους. Επιπλέον, οι ερευνητές έχουν μελετήσει εκτενώς τον εντοπισμό ανεπιθύμητης γνώμης ή ψεύτικων κριτικών σε διαδικτυακά συστήματα αναθεώρησης χρησιμοποιώντας λειτουργίες που βασίζονται στο περιεχόμενο κριτικής και μια ποικιλία μεταδεδομένων.

Συνοψίζοντας, στο παρόν κεφάλαιο εστίασαμε στην πιθανότητα κακής χρήσης μοντέλων βαθιάς μάθησης στο πλαίσιο επίθεσης σε διαδικτυακές πλατφόρμες

αξιολόγησης. Καταδεικνύεται πώς τα RNN μπορούν να δημιουργήσουν παραπλανητικές αλλά ρεαλιστικές κριτικές που στοχεύουν εστιατόρια στο Yelp. Μια εκτενής αξιολόγηση της ποιότητας των δημιουργούμενων κριτικών υποδεικνύει τη δυσκολία στον εντοπισμό τέτοιων κριτικών χρησιμοποιώντας υπάρχουσες αλγοριθμικές προσεγγίσεις, ακόμη και με ανθρώπινη εξέταση (που χρησιμεύει ως δοκιμή από άκρο σε άκρο της επίθεσής μας). Προτείνεται μια νέα προσέγγιση για την άμυνα έναντι ψεύτικων κριτικών που βασίζονται σε RNN, αξιοποιώντας έναν θεμελιώδη περιορισμό ενός μοντέλου που βασίζεται σε RNN: απώλεια πληροφοριών που προκύπτει κατά τη διάρκεια της εκπαιδευτικής διαδικασίας όταν προσαρμόζεται ένα μεγάλο σύνολο δεδομένων εκπαίδευσης σε ένα στατιστικό μοντέλο σταθερού μεγέθους. Λόγω της απώλειας πληροφοριών, οι δημιουργούμενες κριτικές αποκλίνουν από τις πραγματικές κριτικές κατά τη σύγκριση της κατανομής τους σε επίπεδο χαρακτήρων, ακόμη και όταν διατηρούνται γλωσσικά χαρακτηριστικά υψηλότερου επιπέδου. Το σχέδιο που βασίζεται στην εποπτευόμενη μάθηση μπορεί να ανιχνεύσει κριτικές που δημιουργούνται από μηχανή με υψηλή ακρίβεια και ξεπερνά τα υπάρχοντα φίλτρα ψεύτικης κριτικής που βασίζονται σε ML.

Όσον αφορά τις πιθανές μελλοντικές εργασίες, μια κατεύθυνση είναι να εξεταστεί ο ρόλος που μπορούν να διαδραματίσουν τα μεταδεδομένα χρήστη και περιεχομένου τόσο στην προοπτική επίθεσης όσο και στην άμυνα. Τα μεταδεδομένα μπορεί να είναι ζωτικής σημασίας όσον αφορά την εξαπάτηση των χρηστών (π.χ. αυξάνοντας τον αριθμό των φίλων/επαφών στον ιστότοπο) και για την υποστήριξη άμυνας (π.χ. αναλύοντας τα μοτίβα σε χρονικές σημάνσεις των δραστηριοτήτων των χρηστών). Η ενορχήστρωση της γενικής συμπεριφοράς των λογαριασμών χρηστών χρησιμοποιώντας τη βαθιά εκμάθηση για την παράκαμψη άμυνων που βασίζονται σε μεταδεδομένα θα μπορούσε να είναι μια ενδιαφέρουσα ερευνητική πρόκληση. Δεύτερον, ενώ περιοριζόμαστε στον τομέα των διαδικτυακών συστημάτων αναθεώρησης και των επιθέσεων ψεύτικων κριτικών, τα μοντέλα δημιουργίας κειμένου που βασίζονται σε βαθιά μάθηση μπορούν να εφαρμοστούν για την εκτόξευση επιθέσεων και σε άλλα σενάρια. Επισημαίνονται δύο από αυτά τα πιθανά σενάρια εφαρμογής (Fei et al., 2013):

Ενίσχυση των επιθέσεων Sybil: Οι εισβολείς μπορούν να χρησιμοποιήσουν τις τεχνικές για να δημιουργήσουν ρεαλιστικά μοτίβα συμπεριφοράς χρηστών βασισμένα σε κείμενο, π.χ. ανάρτηση, σχολιασμός και ανταλλαγή μηνυμάτων. Αυτό μπορεί να βοηθήσει τους εισβολείς να κάνουν τους (ψεύτικους) λογαριασμούς Sybil να μην διακρίνονται από τους νόμιμους λογαριασμούς που βασίζονται σε περιεχόμενο κειμένου. Μια ειδική περίπτωση αυτού περιλαμβάνει την έναρξη επίθεσης πλαστοπροσωπίας σε διαδικτυακά κοινωνικά δίκτυα.

Fake News Generation: Ο εντοπισμός ψεύτικων ειδήσεων, δηλαδή «μιας φτιαγμένης ιστορίας με πρόθεση εξαπάτησης», παραμένει επί του παρόντος μια ανοιχτή πρόκληση. Η ερευνητική κοινότητα έχει αρχίσει να διερευνά τη δυνατότητα αυτοματοποίησης της διαδικασίας ανίχνευσης με την κατασκευή ενός αγωγού ελέγχου δεδομένων με τη βοήθεια τεχνητής νοημοσύνης (artificial intelligence – AI). Πιστεύεται, ωστόσο, ότι η τεχνητή νοημοσύνη όχι μόνο μπορεί να βοηθήσει στον εντοπισμό ψεύτικων ειδήσεων αλλά και να δημιουργήσει ψεύτικες ειδήσεις. Δεδομένης της διαθεσιμότητας συνόλων δεδομένων ειδήσεων μεγάλης κλίμακας, ένας εισβολέας μπορεί ενδεχομένως να δημιουργήσει ρεαλιστικά άρθρα ειδήσεων χρησιμοποιώντας μια προσέγγιση βαθιάς μάθησης. Λόγω του χαμηλού οικονομικού του κόστους, ο εισβολέας μπορεί να «μολύνει» τις ροές ειδήσεων των μέσων κοινωνικής δικτύωσης με μεγάλο αριθμό ψεύτικων άρθρων.

ΚΕΦΑΛΑΙΟ 6^ο

6.1 Μελέτη περίπτωσης: Amazon

Προκειμένου να προωθήσουμε την κατανόησή μας για το πώς οι άνθρωποι μπορούν να δικαιολογήσουν τη συμμετοχή τους στο crowdurfing, θα αναλύσουμε τη συνεχή μάχη μεταξύ της Amazon, της αμερικανικής εταιρείας ηλεκτρονικού εμπορίου και του τεράστιου αριθμού πλατφορμών crowdurfing. Θα συζητήσουμε στη συνέχεια τους διαφορετικούς ηθικούς λόγους αυτής της συμπεριφοράς μέσω εννοιολογικής ανάλυσης μιας υποδειγματικής πραγματικής υπόθεσης. Συγκεκριμένα, η ανάλυσή μας βασίζεται σε εννοιολογική επιχειρηματολογία σχετικά με την προσαρμογή μεταξύ των

διαφορετικών ηθικών λογικών και των πιθανών συμπεριφορών που προκύπτουν. Τούτου λεχθέντος, πιστεύουμε ότι αυτή η προσέγγιση είναι κατάλληλη για επεξηγηματικούς σκοπούς, ειδικά όταν οι περιπτώσεις που συζητήθηκαν έχουν λάβει σημαντική κάλυψη από τα μέσα ενημέρωσης, επιτρέποντάς μας έτσι να αντλήσουμε σημαντικές γνώσεις (Smith, 2002).

6.2 Περιγραφή υπόθεσης

Η Amazon είναι γνωστή ως ο μεγαλύτερος διαδικτυακός λιανοπωλητής στον κόσμο. Ένα από τα μεγαλύτερα πλεονεκτήματα της Amazon είναι το φιλικό προς το χρήστη σύστημα ανατροφοδότησης, το οποίο επιτρέπει στους χρήστες να αξιολογούν και να υποβάλλουν κριτικές για τα προϊόντα που αγοράζουν στον ιστότοπό της. Στην πραγματικότητα, η Amazon ονομάστηκε η μεγαλύτερη ενιαία πηγή κριτικών στο Διαδίκτυο το 2010. Επομένως, η αξιοπιστία του συστήματος αναθεώρησης είναι πρωταρχικής σημασίας για το επιχειρηματικό μοντέλο της Amazon, καθώς οι αγοραστές του Διαδικτύου εκτιμούν τις αξιόπιστες πληροφορίες σχετικά με τα προϊόντα πριν τα αγοράσουν (Rubin, 2016).

Τα τελευταία χρόνια, τα κυρίαρχα μέσα ενημέρωσης αναφέρουν τη δυσκολία που αντιμετωπίζει η Amazon με τις τοποθεσίες crowdurfing που το περιγράφουν ως ένα «παιχνίδι Whac-a-Mole που δεν τελειώνει ποτέ» (Gonzalez, 2016). Κατά τη διάρκεια αυτής της μάχης, η Amazon έχει ξεχωρίσει και μηνύσει διάφορες πλατφόρμες όπως amazonverifiedreviews.com, amazonreviewsstar.com, buyamazonreviews.info, reviewconnections.com και paybookreviews.org, εκτός από 1.114 άτομα που διαφημίζουν τις υπηρεσίες τους επί πληρωμή στην πλατφόρμα fiverr.com. Ενώ αρκετοί ιστότοποι έχουν κλείσει ως αποτέλεσμα αυτών των νομικών ενεργειών, πολλοί εξακολουθούν να λειτουργούν και νέοι συνεχίζουν να εμφανίζονται.

Μεταξύ αυτών που εξακολουθούν να λειτουργούν τη στιγμή της σύνταξης αυτών των γραμμών είναι το paybookreviews.org, το οποίο χρησιμοποιείται ως ενδεικτικό παράδειγμα. Στο εμπορικό της βίντεο διάρκειας 37 δευτερολέπτων, οι Κριτικές βιβλίων επί πληρωμή (Paid Book Reviews - PBR) διαφημίζονται ως εξής: «Οι κριτικές βιβλίων

επί πληρωμή - η dot org είναι μια ομάδα συγγραφέων που κατανοούν την επίδραση των θετικών κριτικών πελατών στις πωλήσεις των βιβλίων σας. Προσφέρουμε δύο τύπους κριτικών βιβλίων: οι πρώτοι είναι μη επαληθευμένες κριτικές βιβλίων. Διαβάζουμε τα δείγματα σελίδων του βιβλίου σας στο Amazon dot com και στη συνέχεια, δημοσιεύουμε θετικά σχόλια κάτω από τις κριτικές πελατών. Το δεύτερο είναι επαληθευμένες κριτικές βιβλίων. Αγοράζουμε το βιβλίο σας και το διαβάζουμε στο Kindle και μετά δημοσιεύουμε θετικά σχόλια για το βιβλίο σας στο Amazon. Επισκεφτείτε το www.payedbookreviews.org τώρα!» (Rinta-Kahila & Soliman, 2017).

Ως εκ τούτου, η PBR υπόσχεται στους συνδρομητές της θετικές κριτικές για τα βιβλία τους που αναφέρονται στο Amazon. Για παράδειγμα, ο ιστότοπος προωθεί ότι στους συγγραφείς που επιλέγουν την επιλογή «επαληθευμένες κριτικές βιβλίων» υπόσχονται 50 θετικές κριτικές για ένα σύνολο 1250 \$. Ενώ η εταιρεία ισχυρίζεται ότι οι εργαζόμενοί της αγοράζουν και διαβάζουν πραγματικά τα βιβλία που τους έχουν ανατεθεί, η Amazon δεν είναι τόσο πεπεισμένη καθώς ζήτησε από τον ιστότοπο να σταματήσει να χρησιμοποιεί το εμπορικό σήμα της Amazon και να σταματήσει να προσφέρει κριτικές για την Amazon προς πώληση. Η Amazon έχει δηλώσει συγκεκριμένα στις Οδηγίες κοινότητας, ότι απαγορεύει ρητά τη δημιουργία περιεχομένου σε αντάλλαγμα για αποζημίωση ή για λογαριασμό οποιουδήποτε άλλου. Επιπλέον, οι οδηγίες δεν επιτρέπουν την προσφορά αποζημίωσης για τέτοια παραγωγή περιεχομένου. Σημειώστε ότι αυτή η υποδειγματική περίπτωση crowdurfing περιλαμβάνει τρεις παράγοντες που εμπλέκονται άμεσα στη σύνταξη κριτικών επί πληρωμή: τους πελάτες που χρειάζονται (και πληρώνουν για) τις θετικές κριτικές, τους εργαζόμενους, οι οποίοι προσλαμβάνονται για να γράψουν τις θετικές κριτικές βιβλίων με αντάλλαγμα μια αμοιβή και τον πράκτορα που ενορχηστρώνει την πλατφόρμα και διαφημίζεται στους πελάτες και προσκαλεί τους εργαζόμενους να γράψουν τις κριτικές. Η συμπεριφορά αυτών των παραγόντων επηρεάζει άλλα ενδιαφερόμενα μέρη που δεν εμπλέκονται άμεσα στη σύνταξη των αξιολογήσεων - κυρίως α) την πλατφόρμα-στόχο της Amazon της οποίας η αξιοπιστία του συστήματος αξιολόγησης επηρεάζεται από τις πληρωμένες κριτικές και β) την ευρύτερη κοινότητα χρηστών της οποίας τα μέλη αναζητούν και χρησιμοποιούν τις πληροφορίες που είναι διαθέσιμες στην πλατφόρμα-στόχο.

6.3 Ανάλυση περίπτωσης

Σε αυτήν την ενότητα, συζητάμε αυτό το εν συντομία παράδειγμα crowdfunding από διαφορετικές ηθικές προοπτικές για κάθε εμπλεκόμενο παράγοντα που εντοπίζεται σε ένα από τα τρία ηθικά επιχειρήματα: παραδοσιακή φιλοσοφική ηθική, επιχειρηματική ηθική και κωδικοποιημένοι κανόνες. Η αρχή που ενσωματώνεται σε μια συγκεκριμένη ηθική προοπτική μπορεί να θεωρηθεί ότι συνεπάγεται μια κατευθυντήρια γραμμή συμπεριφοράς για κάθε δεδομένη ενέργεια.

6.3.1 Οι πελάτες

Οι πελάτες που αγοράζουν κατασκευασμένες κριτικές για τα βιβλία τους (ή των πελατών τους) μέσω PBR αναμένουν ότι οι κριτικές θα οδηγήσουν σε θετικά οικονομικά αποτελέσματα που μπορεί να προκύψουν, για παράδειγμα, από αυξημένες πωλήσεις που αποδίδονται σε θετικές κριτικές για το δικό τους προϊόν ή, σε ορισμένες περιπτώσεις, σε αρνητικές αξιολογήσεις των προϊόντων των ανταγωνιστών. Ίσως το πιο σημαντικό, ωστόσο, είναι πως οι πελάτες έχουν μεγαλύτερη προβολή για τα βιβλία τους. Στην πραγματικότητα, πρόσφατη έρευνα IS δείχνει ότι οι ψεύτικες κριτικές θα μπορούσαν πράγματι να αυξήσουν την προβολή στο διαδίκτυο. Στη σημερινή υπερχειλίση πληροφοριών που βιώνουμε, μπορεί να είναι εμφανώς δύσκολο να ξεχωρίσεις από τη μάζα, επομένως η αγορά κριτικών μπορεί να φαίνεται ως ο μόνος δυνατός τρόπος για να τραβήξει την προσοχή ο συγγραφέας. Έτσι, από την τελεολογική/συνεπακόλουθη σκοπιά, η αγορά κατασκευασμένων κριτικών θα μπορούσε να θεωρηθεί αποδεκτή καθώς συνεπάγεται πολλά θετικά αποτελέσματα: οι πελάτες επωφελούνται από υψηλότερη προβολή και πωλήσεις, οι εργαζόμενοι επωφελούνται από τον καταβαλλόμενο μισθό και οφέλη PBR από την αύξηση των εργασιών (Lappas et al., 2016).

Επιπλέον, ο πελάτης μπορεί να θεωρήσει ότι το βιβλίο του έχει θετικό αντίκτυπο στον αναγνώστη και επομένως η πώληση περισσότερων από αυτό θα ήταν επιθυμητός στόχος. Η δεοντολογική προοπτική, ωστόσο, υπενθυμίζει ότι το να πουλά κανείς τα

προϊόντα του πληρώνοντας άλλους για να λένε ψέματα για αυτά είναι ηθικά λάθος και ως εκ τούτου δεν είναι αποδεκτό. Υποθέτοντας την οπτική των μετόχων, θα μπορούσε να συναχθεί το συμπέρασμα ότι οι αξιολογήσεις αγοράς μπορεί να θεωρηθούν αποδεκτές, καθώς τόσο οι πελάτες όσο και η PBR επωφελούνται από την άμεση συναλλαγή. Ωστόσο, τέτοιες ενέργειες δεν υφίστανται κριτική αξιολόγηση από την πλευρά των ενδιαφερομένων, λόγω του ίδιου συλλογισμού που συζητήθηκε προηγουμένως. Συγκεκριμένα, ενώ οι πελάτες έχουν κίνητρα να μεγιστοποιήσουν τα δικά τους κέρδη, κάνουν ελάχιστα για να ελαχιστοποιήσουν τα μειονεκτήματα για την Amazon, την κοινότητα του Διαδικτύου και τους ανταγωνιστές τους των οποίων οι πωλήσεις ενδέχεται να υποφέρουν άδικα λόγω δυνητικά αναληθών κριτικών. Τέλος, ενώ η λειτουργία τους μπορεί να είναι νομικά έγκυρη, οι πελάτες δεν τηρούν τις Οδηγίες κοινότητας της Amazon που απαγορεύουν την προσφορά αποζημίωσης για τη δημιουργία, την τροποποίηση ή τη δημοσίευση περιεχομένου. Αυτές οι οδηγίες μπορούν να θεωρηθούν ως κωδικοποιημένοι κανόνες που έχουν σχεδιαστεί για να μετριάζουν τη συμπεριφορά και την αλληλεπίδραση των μελών της κοινότητας της Amazon στον ιστότοπο (Lappas et al., 2016).

6.3.2 Εργαζόμενοι

Οι εργαζόμενοι που γράφουν κατασκευασμένες κριτικές κατόπιν ζήτησης λαμβάνουν οικονομική αποζημίωση για τις προσπάθειές τους, ενώ δεν αντιμετωπίζουν κίνδυνο κυρώσεων. Έτσι, μια τέτοια κατάσταση αποτελεί πρόσφορο έδαφος για ευκαιριακή συμπεριφορά, την οποία οι εργαζόμενοι μπορούν να δικαιολογήσουν χρησιμοποιώντας τελεολογική συλλογιστική. Αν και οι εργαζόμενοι μπορεί να μην είναι ειλικρινείς στις κριτικές τους, μπορεί να υποστηρίξουν ότι αυτό πρέπει να κάνουν για να βγάλουν τα προς το ζην. Επιπλέον, μπορεί να πείσουν τους εαυτούς τους ότι οι κριτικές θα οδηγούσαν σε υψηλότερο συνολικό κέρδος καθώς οι συνέπειες είναι θετικές για αρκετούς βασικούς παράγοντες. Ο εργαζόμενος αμείβεται, ο πελάτης λαμβάνει την υπηρεσία και το PBR παίρνει δουλειά. Αν και η ανεντιμότητα των κριτικών επηρεάζει αρνητικά την αξιοπιστία της Amazon και τη συνολική ποιότητα των πληροφοριών στο Διαδίκτυο, οι εργαζόμενοι μπορεί να σκεφτούν ότι είναι απίθανο μερικές πληρωμένες κριτικές να κάνουν τη διαφορά. Τα πιθανά θύματα είναι πιθανώς σωματικά πολύ μακριά και, επομένως, μπορεί να φαίνονται εικονικά ή ανύπαρκτα για τους turfers. Ωστόσο, εάν κάποιος επρόκειτο να χρησιμοποιήσει δεοντολογικό συλλογισμό, η

σύνταξη κατασκευασμένων κριτικών θα θεωρούνταν απαράδεκτη, καθώς το ψέμα είναι κατηγορηματικά λάθος και επομένως δεν είναι δικαιολογημένο μέσο, ανεξάρτητα από το πόσο επιθυμητό μπορεί να είναι το τελικό αποτέλεσμα (Chatterjee et al. 2015).

Ενώ η προοπτική των μετόχων μπορεί να μην είναι η πιο σχετική λογική από την οπτική γωνία ενός εργαζομένου, η σύνταξη κατασκευασμένων κριτικών μπορεί να θεωρηθεί ότι ωφελεί τους κύριους μετόχους που εμπλέκονται στο crowdfunding, δηλαδή τους πελάτες και το PBR. Ωστόσο, η κατάσταση φαίνεται διαφορετική όταν λαμβάνονται υπόψη άλλοι ενδιαφερόμενοι: η αξιοπιστία της Amazon υποφέρει και η ποιότητα των πληροφοριών που είναι διαθέσιμες σε άλλους χρήστες του Διαδικτύου υποβαθμίζεται. Ομοίως, ο φακός της θεωρίας των κοινωνικών συμβολαίων αποκαλύπτει ότι ενώ οι εργαζόμενοι μεγιστοποιούν τα δικά τους πλεονεκτήματα με το crowdfunding, δεν ελαχιστοποιούν ταυτόχρονα τα μειονεκτήματα για την Amazon και την κοινότητα του Διαδικτύου. Τέλος, οι Οδηγίες κοινότητας της Amazon αναφέρουν ότι «δεν επιτρέπεται η δημιουργία, τροποποίηση ή δημοσίευση περιεχομένου με αντάλλαγμα αποζημίωση οποιουδήποτε είδους (συμπεριλαμβανομένων δωρεάν προϊόντων ή προϊόντων με έκπτωση)». Επομένως, αυτοί οι κωδικοποιημένοι κανόνες απαγορεύουν ρητά στους κριτικούς να γράφουν κριτικές με αντάλλαγμα την αποζημίωση. Ωστόσο, η σύνταξη ενός μη αυθεντικού σχολίου σε ένα προϊόν δεν φαίνεται να έχει σημασία για την επίκληση της επιβολής του νόμου (Chatterjee et al. 2015).

6.3.3 Μέσο

Από τη σκοπιά του αντιπροσώπου (σε αυτήν την περίπτωση, PBR), η σύνδεση εργαζομένων με πελάτες μπορεί να θεωρηθεί ότι έχει θετικές οικονομικές συνέπειες, καθώς η πλατφόρμα τους παρέχει στους εργαζόμενους την ευκαιρία να κερδίσουν χρήματα και στους πελάτες πρόσβαση σε προσιτή εργασία. Έτσι, η παροχή τέτοιων υπηρεσιών θα μπορούσε να θεωρηθεί αποδεκτή από τελεολογική άποψη, δεδομένου ότι υπάρχουν θετικές συνέπειες για πολλά ενδιαφερόμενα μέρη. Ωστόσο, από δεοντολογική άποψη, η προσέλκυση και η ενθάρρυνση των εργαζομένων να λένε ψέματα και οι πελάτες να πληρώνουν για ψέματα είναι κατηγορηματικά λάθος και ως εκ τούτου απαράδεκτη. Από την πλευρά ενός μετόχου, η δραστηριότητα της PBR δεν είναι προβληματική, καθώς ωφελούνται τόσο οι μέτοχοι της PBR όσο και οι μέτοχοι των πελατών τους. Και πάλι, η PBR δεν λαμβάνει υπόψη τις ανάγκες άλλων

ενδιαφερόμενων μερών, δηλαδή της Amazon και της κοινότητας του Διαδικτύου, καθώς η υπηρεσία της διεγείρει συμπεριφορά που βλάπτει και τα δύο. Επιπλέον, από την άποψη του κοινωνικού συμβολαίου, το PBR μεγιστοποιεί μόνο τα δικά του πλεονεκτήματα και τα πλεονεκτήματα των πελατών του. Όχι μόνο δεν κάνει τίποτα για να ελαχιστοποιήσει τα μειονεκτήματα για την Amazon και την κοινότητα του Διαδικτύου από τις ενέργειές τους, αλλά στην πραγματικότητα αυξάνει αυτά τα μειονεκτήματα: όσο περισσότεροι συντάκτες κριτικών καταφέρνουν να συνδεθούν με πελάτες που πωλούν προϊόντα στο Amazon, τόσο πιο αναξιόπιστο γίνεται το σύστημα αξιολόγησης της Amazon και τόσο λιγότερο αξιόπιστες πληροφορίες θα είναι διαθέσιμες για τους χρήστες του Διαδικτύου. Τέλος, δεδομένου ότι το PBR δεν λειτουργεί στην ουσία απευθείας στην ηλεκτρονική πλατφόρμα της Amazon, δεν παραβιάζει άμεσα τους κωδικοποιημένους κανόνες της Amazon. Ωστόσο, καλεί και ενθαρρύνει άλλους παράγοντες να τις παραβιάσουν, προωθώντας δόλιες δραστηριότητες (Rinta-Kahila & Soliman, 2017).

Συζήτηση

Σύμφωνα με τον Smith (2002), αναγνωρίζουμε ότι ένας τυπικός χρήστης IS μπορεί να βασίσει την ηθική του επιχειρηματολογία σε διάφορες λογικές μέσα στις προοπτικές που συζητήθηκαν προηγουμένως, αλλά συνήθως μια προοπτική είναι ο κύριος μοχλός της επιχειρηματολογίας των χρηστών. Με συνέπεια, η ανάλυσή μας για μια περίπτωση crowdurfing καταδεικνύει ότι κάθε βασικός παράγοντας που εμπλέκεται στο crowdurfing μπορεί να δικαιολογήσει την ηθική συμπεριφοράς του στηριζόμενη σε μία από τις θεωρίες ή τα επίπεδα κανόνων σε έναν δεδομένο τομέα. Δεδομένου ότι διαφορετικές ηθικές θεωρίες ή κανόνες τείνουν να παρέχουν αντικρουόμενες ηθικές

κατευθυντήριες γραμμές, ένας άνθρωπος πρέπει συνήθως να απορρίπτει εναλλακτικές απόψεις όταν συμμορφώνεται με την επιλεγμένη άποψη.

Οι Chatterjee et al. (2015) σημειώνουν ότι ο ομορτισμός είναι κάτι πολύ εγγενές σε όλους τους ανθρώπους και είναι πιθανό ότι οι φορείς που εμπλέκονται στο crowdfunding παρακινούνται κυρίως να μεγιστοποιήσουν το προσωπικό τους συμφέρον απουσία υφιστάμενων κυρώσεων. Συνολικά, κάθε παράγοντας μπορεί να χρησιμοποιήσει μια σειρά τεχνικών εξουδετέρωσης που βασίζονται σε ορισμένες ηθικές προοπτικές. Στον τομέα της παραδοσιακής φιλοσοφικής ηθικής, φαίνεται ότι οι ηθοποιοί που εμπλέκονται στο crowdfunding στηρίζονται στην τελεολογική άποψη ενώ απορρίπτουν τη δεοντολογική άποψη. Πράγματι, η συμβολή στην παραγωγή ψευδών πληροφοριών είναι δύσκολο να δικαιολογηθεί από δεοντολογική άποψη, η οποία τείνει να συνεπάγεται κατηγορηματικές επιταγές όπως «δεν πρέπει να λες ψέματα». Αν αυτή ήταν η ηθική κατευθυντήρια γραμμή κάθε χρήστη του Διαδικτύου, η επιχείρηση crowdfunding δεν θα μπορούσε να υπάρξει.

Ωστόσο, οι εμπλεκόμενοι φορείς μπορούν να δικαιολογήσουν τη συμπεριφορά τους προσκολλόμενοι σε μια υποκειμενική ερμηνεία της τελεολογικής προοπτικής, αν και με ορισμένους περιορισμούς. Συγκεκριμένα, εστιάζοντας αυστηρά στις βραχυπρόθεσμες οικονομικές συνέπειες για τους παράγοντες που εμπλέκονται άμεσα στην εκστρατεία crowdfunding, θα μπορούσε κανείς να υποστηρίξει ότι από τελεολογική άποψη το crowdfunding είναι ηθικό. Επιπλέον, ένας πελάτης μπορεί συχνά να παρακινείται από άλλες εκτός από οικονομικές φιλοδοξίες, όπως η απόκτηση προβολής και αναγνώρισης ως συγγραφέας (Siponen et al., 2012).

Όταν εξετάζουμε την προσέγγιση της επιχειρηματικής ηθικής, είναι πιθανό οι φορείς που εμπλέκονται στο crowdfunding να βασίζονται στη θεωρία των μετόχων, ενώ θα απορρίπτουν τις θεωρίες των ενδιαφερομένων και των κοινωνικών συμβολαίων. Η αυστηρή τήρηση του δόγματος του Friedman που υποχρεώνει κάποιον να επικεντρωθεί στη μεγιστοποίηση του πλεονεκτήματος των μετόχων είναι ένα απλό επιχείρημα υπέρ του crowdfunding. Οι υποστηρικτές αυτής της άποψης δεν χρειάζεται να ασχοληθούν με το να λάβουν υπόψη τα συμφέροντα άλλων ενδιαφερομένων ή να ελαχιστοποιήσουν τις ζημιές σε άλλα μέρη που προκαλούνται από τη δραστηριότητά τους (Rinta-Kahila & Soliman, 2017).

Στην πραγματικότητα, εάν εξετάζαν αυτά τα ζητήματα, το crowdfunding θα ήταν αδύνατο να δικαιολογηθεί ηθικά. Επιπλέον, όσον αφορά τη θεωρία του κοινωνικού συμβολαίου, οι φορείς που εμπλέκονται στο crowdfunding μπορεί να αντιληφθούν ότι οι ενέργειες στον κυβερνοχώρο διαφέρουν από τις ενέργειες στον πραγματικό κόσμο και επομένως δεν αισθάνονται υποχρεωμένοι να ενεργούν σύμφωνα με το ίδιο κοινωνικό συμβόλαιο ή κοινωνικούς κανόνες όταν είναι διαδικτυακοί. Για αυτούς, ό,τι δεν είναι αποδεκτό στον φυσικό κόσμο, μπορεί να είναι αποδεκτό στον εικονικό κόσμο. Αν και η θεωρία των μετόχων υποχρεώνει τον δρώντα να παραμείνει «εντός των κανόνων του παιχνιδιού», οι φορείς μπορούν να επιλέξουν να αντλήσουν τους κανόνες από το επίπεδο κωδικοποιημένων κανόνων που τους ταιριάζει καλύτερα, π.χ. από αποτελεσματική νομοθεσία αντί από τους όρους και τις προϋποθέσεις της πλατφόρμας-στόχου. Έτσι, το επίπεδο των κωδικοποιημένων κανόνων που λαμβάνονται υπόψη μπορεί να καθορίσει τους ηθικούς λόγους του crowdfunding (Rinta-Kahila & Soliman, 2017).

Η πώληση, η αγορά ή η διαμεσολάβηση κατασκευασμένης υποστήριξης, όπως ψεύτικες κριτικές, δεν είναι απαραίτητως παράνομη, επομένως κανένας από τους παράγοντες δεν χρειάζεται να ασχοληθεί με νομικά προβλήματα. Έτσι, φαίνεται ότι κρίνουν αποδεκτή τη συμπεριφορά τουλάχιστον εφόσον δεν παραβιάζει κανόνες σε επίπεδο γενικού νόμου. Παρόλο που κωδικοποιημένοι κανόνες υπάρχουν συνήθως σε επίπεδο πλατφόρμας (όπως η Amazon), αυτές οι οδηγίες ενδέχεται να μην έχουν σημαντικό αντίκτυπο στους φορείς. Αυτό σχετίζεται με ένα κοινό πρόβλημα στις υπηρεσίες λογισμικού, καθώς τέτοιες οδηγίες τείνουν να απορρίπτονται εύκολα ως ασήμαντες πληροφορίες που δεν έχουν καμία πρακτική συνάφεια. Για παράδειγμα, διαπιστώθηκε ότι μόνο το 7% των ανθρώπων διαβάζουν τους όρους και τις προϋποθέσεις όταν αγοράζουν ένα προϊόν ή μια υπηρεσία στο διαδίκτυο (Smithers, 2011).

Σύμφωνα με τους Willison & Warkentin (2013), η ανάλυσή τους υπογραμμίζει τη σημασία της κατανόησης των ηθικών ριζών και κινήτρων πίσω από τη συμπεριφορά που απειλεί την εγκυρότητα των πληροφοριών και μπορεί έτσι να γίνει κίνδυνος ασφάλειας στην εποχή της πληροφορίας. Επιπλέον, παρέχουν μια πιο πολύπλευρη οπτική του φαινομένου: κοιτάζοντας διαφορετικά τους τρεις βασικούς παράγοντες που

εμπλέκονται άμεσα στο crowdurfing, διαπιστώνουν ότι - αν και νωρίτερα τους παρουσίασαν ως ξεχωριστές οντότητες από τον υπόλοιπο πληθυσμό του Διαδικτύου – είναι επίσης χρήστες του Διαδικτύου που απολαμβάνουν τα πλεονεκτήματα των ελεύθερων διαθέσιμων πληροφοριών.

Έτσι, ενώ η συμβολή στην ύπαρξη εκστρατειών crowdurfing δημιουργεί οικονομικές ανταμοιβές για τα εμπλεκόμενα μέρη, μειώνει την αξία των πληροφοριών που είναι διαθέσιμες για όλους τους χρήστες του Διαδικτύου, συμπεριλαμβανομένων των ίδιων. Αυτή η αντίφαση θα μπορούσε να εξηγηθεί από την αμέλεια των ηθοποιών για τη θέση της καθολικότητας, δηλαδή, με την ερώτηση «τι θα γινόταν αν το έκαναν όλοι;» Εάν κάθε χρήστης του Διαδικτύου είτε αγόραζε είτε παρήγαγε ψεύτικο περιεχόμενο είτε ενεργούσε ως ενδιάμεσος μεταξύ των δύο, σύντομα δεν θα υπήρχαν αξιόπιστες πληροφορίες διαθέσιμες στον κοινόχρηστο κυβερνοχώρο (Rinta-Kahila & Soliman, 2017).

Επιπλέον, δεν θα υπήρχε λογική απαίτηση για πρόσθετους «ακόλουθους», «μου αρέσει» ή «κριτικές», εάν η προεπιλεγμένη υπόθεση ήταν ότι δεν είναι όλοι αξιόπιστοι. Ωστόσο, ειδικά οι εργαζόμενοι μπορεί να μην πιστεύουν ότι μια μικρής κλίμακας παραγωγή ψευδούς περιεχομένου έχει σημασία σε μεγάλη κλίμακα, καθώς θα μπορούσαν να υποθέσουν ότι οι περισσότεροι χρήστες του Διαδικτύου είναι ακόμα ειλικρινείς (Rinta-Kahila & Soliman, 2017).

Συμπέρασμα

Το crowdurfing, που επινοήθηκε από τους Wang et al. (2012), συνδυάζει το crowdsourcing με το astroturfing. Το τελευταίο αναφέρεται σε μια πρακτική όπου ένα συγκεκριμένο μέρος (άτομο ή οργανισμός) επιχειρεί να διαδώσει πληροφορίες και να δημιουργήσει συναίνεση για ένα θέμα του ενδιαφέροντός του με τρόπο που μιμείται δραστηριότητες γνήσιων κινημάτων βάσης. Αυτό συμβαίνει συχνά με τη μορφή διάδοσης συγκεκριμένων φημών, πολιτικών μηνυμάτων ή ψευδών διαφημίσεων. Το

crowdturfing είναι μια συγκεκριμένη μορφή astroturf, όπου ο πελάτης αξιοποιεί τη δύναμη του πλήθους κινητοποιώντας τους ανθρώπους μέσω του Διαδικτύου για να συμμετάσχουν σε καμπάνιες που προάγουν τον στόχο του πελάτη. Τέτοιες εκστρατείες συνήθως δομούνται ως συλλογές εργασιών που ολοκληρώνουν οι εργαζόμενοι με αντάλλαγμα μια αμοιβή.

Σε αντίθεση με το astroturfing, η γρήγορη και αποτελεσματική χρήση πραγματικών ανθρώπων εργαζομένων θεωρείται ότι αυξάνει δραστικά τον αντίκτυπο τέτοιων εκστρατειών. Ενώ πολλές πλατφόρμες crowdsourcing απαγορεύουν ρητά και προσπαθούν να ανακόψουν τέτοιες εκστρατείες, το crowdturfing καταλήγει συχνά ως η βασική δραστηριότητα τέτοιων ιστότοπων (π.χ. Αγορά Likes & Followers).

Είναι σημαντικό να σημειωθεί ότι αυτή η μελέτη είναι βιβλιογραφικής φύσης και ως εκ τούτου, είναι μόνο ένα πρώτο βήμα προς την κατανόηση της συμπεριφοράς. Ένα φυσικό επόμενο βήμα θα ήταν να βασιστούμε στις γνώσεις που έχουμε για να προχωρήσουμε με τη διεξαγωγή εμπειρικής εργασίας (π.χ. με τη διεξαγωγή ποιοτικών και ποσοτικών μελετών). Ωστόσο, αναγνωρίζουμε ότι η απόκτηση πρόσβασης σε εμπειρικά δεδομένα απευθείας από φορείς που εμπλέκονται σε αυτόν τον επιχειρηματικό τομέα μπορεί να είναι πρόκληση για τους μελλοντικούς ερευνητές. Ενώ αυτές οι επιχειρήσεις μπορεί να λειτουργούν στο πλαίσιο του κράτους δικαίου, οι ηθικοί τους λόγοι μπορεί να είναι σαθροί, γεγονός που θα μπορούσε να αποθαρρύνει τους φορείς από το να συνεργαστούν με ερευνητές. Ως εκ τούτου, οι μελλοντικοί ερευνητές μπορούν να χρησιμοποιήσουν ορισμένες στρατηγικές που θα μπορούσαν να ενθαρρύνουν τους εργαζόμενους στο crowdturfing να συνεργαστούν σε εμπειρική έρευνα, π.χ. χρησιμοποιώντας έρευνες βασισμένες σε σενάρια που παρουσιάζουν υποθετικές καταστάσεις με τη μορφή αφηγηματικών βινιετών. Επίσης, η πληρωμή των εργαζομένων για τη συμμετοχή τους στη συλλογή δεδομένων πιθανότατα θα τους έδινε περαιτέρω κίνητρα για συνεργασία. Επιπλέον, οι μελλοντικοί ερευνητές θα μπορούσαν να εξετάσουν την προοπτική των επηρεαζόμενων ενδιαφερομένων μελετώντας, π.χ., πώς οι πλατφόρμες των θυμάτων αντιδρούν και ανταποκρίνονται στις εκστρατείες crowdturfing.

Επιπλέον, ενθαρρύνουμε τη μελλοντική διερεύνηση προληπτικών μέτρων. Ακολουθώντας το παράδειγμα της περίπτωσης ανεπιθύμητης αλληλογραφίας, η

παρακολούθηση της ροής των χρημάτων για την εύρεση της οικονομικής συμφόρησης των εκστρατειών crowdfunding θα μπορούσε να είναι μια προσέγγιση για την αντιμετώπιση του προβλήματος. Δεδομένου ότι τα οικονομικά κέρδη είναι πιθανό να κατέχουν σημαντική θέση στην παρακίνηση της συμπεριφοράς crowdfunding, η μειωμένη ικανότητα των πλατφορμών crowdfunding να παρέχουν αποζημίωση στους εργαζομένους τους θα μπορούσε να μειώσει σημαντικά την ελκυστικότητά τους. Είναι προφανές ότι η έρευνα για την ανίχνευση του crowdfunding βρίσκεται σε πλήρη εξέλιξη και απαιτούνται πιο αποτελεσματικές μέθοδοι. Μια τέτοια έρευνα θα μπορούσε ενδεχομένως να παρέχει σε πλατφόρμες-στόχους συγκεκριμένα εργαλεία για την αντιμετώπιση των επιβλαβών επιπτώσεων με διορθωτικές ενέργειες, όπως η επισήμανση και η εξάλειψη του ψευδούς περιεχομένου.

Τέλος, αναγνωρίζουμε ότι η θεωρία της αποτροπής είναι η κυρίαρχη προσέγγιση που ενημερώνει τα πρότυπα ασφάλειας πληροφοριών και τις συμπεριφορές ασφάλειας πληροφοριών γενικά. Ωστόσο, θα πρέπει να προσέχουμε ότι η αποτελεσματικότητά της είναι μάλλον περιορισμένη απουσία αυστηρών κυρώσεων και το πιο σημαντικό όταν τα άτομα εφαρμόζουν τακτικές εξουδετέρωσης. Έτσι, προτείνουμε ότι μια βαθύτερη κατανόηση των γεγονότων που προηγούνται της συμπεριφοράς του crowdfunding είναι ένα καλό σημείο εκκίνησης στην αναζήτηση μέσων για την αποτροπή του.

Βιβλιογραφία

Anderson, H. S., Woodbridge, J., & Filar, B. (2016, October). DeepDGA: Adversarially-tuned domain generation and detection. In *Proceedings of the 2016 ACM workshop on artificial intelligence and security* (pp. 13-21).

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management science*, 57(8), 1485-1509.

Berg, L., Sletteameås, D., Kjørstad, I., & Rosenberg, T. G. (2020). Trust and the don't-want-to-complain bias in peer-to-peer platform markets. *International Journal of Consumer Studies*, 44(3), 220-231.

Boukef, N., & Charki, M. (2014). When the dark side of post-adoptive use leads to IT discontinuance: An exploration of the role of intervention.

Chatterjee, S., Sarker, S., & Valacich, J. S. (2015). The behavioral roots of information systems security: Exploring key factors related to unethical IT use. *Journal of Management Information Systems*, 31(4), 49-87.

De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6), 817-833.

Fang, J., Hu, L., Hossin, M. A., Yang, J., & Shao, Y. (2019). Polluted online reviews: The effect of air pollution on reviewer behavior. *International Journal of Electronic Commerce*, 23(4), 557-594.

Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 1, pp. 175-184).

Givel, M. (2007). Consent and counter-mobilization: the case of the National Smokers Alliance. *Journal of health communication*, 12(4), 339-357.

González, Á. (2016). Amazon sues alleged providers of fake reviews. *The Seattle Times, News*, 25.

Gordon, M. E., & De Lima-Turner, K. (1997). Consumer attitudes towards Internet advertising: A social contract perspective. *International Marketing Review*.

Han, R. (2015). Manufacturing consent in cyberspace: China's "fifty-cent army". *Journal of Current Chinese Affairs*, 44(2), 105-134.

- Henry, R. (2015). How to fake a bestseller. *The Sunday Times*.
- Ho, S. M., & Hancock, J. T. (2019). Context in a bottle: Language-action cues in spontaneous computer-mediated deception. *Computers in Human Behavior, 91*, 33-41.
- Hovy, D. (2016, August). The enemy in your own camp: How well can we detect statistically-generated fake reviews—an adversarial study. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 351-356).
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.
- Kaur, R., & Singla, D. S. (2019). Performance measurement of higher educational institutions: An empirical study using student's perception. *Journal of Management, 6*(2).
- Kumar, S., & Purbey, S. (2018). Benchmarking model for factors influencing creation of negative electronic word of mouth. *Benchmarking: An International Journal*.
- Kusuma, L., Rejeki, S., Robiyanto, R., & Irviana, L. (2020). Reputation system of C2C e-commerce, buying interest and trust. *Business: Theory and Practice, 21*(1), 314-321.
- Lappas, T. (2012, June). Fake reviews: The malicious perspective. In *International Conference on Application of Natural Language to Information Systems* (pp. 23-34). Springer, Berlin, Heidelberg.
- Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research, 27*(4), 940-961.
- Lebek, B., Uffen, J., Neumann, M., Hohler, B., & Breitner, M. H. (2014). Information security awareness and behavior: a theory-based literature review. *Management Research Review*.
- Lee, K., Webb, S., & Ge, H. (2014, May). The dark side of micro-task marketplaces: Characterizing fiverr and automatically detecting crowdturfing. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- Βασίλειος Χατζηζήσης, Πανεπιστήμιο Αιγαίου, Τμήμα Μηχ/κων ΠΕΣ*

Lee, K., Webb, S., & Ge, H. (2015). Characterizing and automatically detecting crowdturfing in Fiverr and Twitter. *Social Network Analysis and Mining*, 5(1), 1-16.

Leetaru, K. (2016). Is The Internet Evolving Away From Freedom of Speech?.

Levchenko, K., Pitsillidis, A., Chachra, N., Enright, B., Félegyházi, M., Grier, C., ... & Savage, S. (2011, May). Click trajectories: End-to-end analysis of the spam value chain. In *2011 IEEE Symposium on Security and Privacy* (pp. 431-446). IEEE.

Lim, S. H., Kim, S. W., Kim, S., & Park, S. (2011, March). Construction of a blog network based on information diffusion. In *Proceedings of the 2011 ACM symposium on applied computing* (pp. 937-941).

Lipton, Z. C., Vikram, S., & McAuley, J. (2015). Generative concatenative nets jointly learn to write and classify reviews. *arXiv preprint arXiv:1511.03683*.

Loch, K. D., Carr, H. H., & Warkentin, M. E. (1992). Threats to information systems: today's reality, yesterday's understanding. *Mis Quarterly*, 173-186.

Marinescu, I., Klein, N., Chamberlain, A., & Smart, M. (2018). *Incentives can reduce bias in online reviews* (No. w24372). National Bureau of Economic Research.

Mason, O. R. (1986). Four ethical issues of the information age. Retrieved June 28, 2007.

Mason, O. R. (1986). Four ethical issues of the information age. Retrieved June 28, 2007.

Menn, J., & Sanders, E. (2008). Lobbyists tied to microsoft wrote citizens' letters. *The LA Times* (August 23, 2001).

Plotkina, D., Munzel, A., & Pallud, J. (2020). Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research*, 109, 511-523.

Poushneh, Atieh, and Reza Rajabi. "Can reviews predict reviewers' numerical ratings? The underlying mechanisms of customers' decisions to rate products using Latent Dirichlet Allocation (LDA)." *Journal of Consumer Marketing* (2022).

Price, S. R. (2014). Big Business: Buying Fake Instagram Followers. *The Huffington Post*.

Qahri-Saremi, H., & Montazemi, A. R. (2019). Factors affecting the adoption of an electronic word of mouth message: A meta-analysis. *Journal of Management Information Systems*, 36(3), 969-1001.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2010). Detecting and tracking the spread of astroturf memes in microblog streams. *arXiv preprint arXiv:1011.3768*.

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011, March). Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World wide web* (pp. 249-252).

Rayana, S., & Akoglu, L. (2015, August). Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining* (pp. 985-994).

Rinta-Kahila, T., & Soliman, W. (2017). Understanding crowdturfing: the different ethical logics behind the clandestine industry of deception. In *ECIS 2017: Proceedings of the 25th European Conference on Information Systems, Guimarães, Portugal, June 5-10, 2017, ISBN 978-989-20-7655-3*. European Conference on Information Systems.

Roberts, J. (2015). Amazon sues people who charge \$5 for fake reviews. *Fortune Magazine*.

Robson, K., Farshid, M., Bredican, J., & Humphrey, S. (2013). Making sense of online consumer reviews: A methodology. *International Journal of Market Research*, 55(4), 521-537.

Rubin, B. F. (2016). Amazon continues crackdown on alleged fake reviews. *CNet.com*.

Schleimer, S., Wilkerson, D. S., & Aiken, A. (2003, June). Winnowing: local algorithms for document fingerprinting. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data* (pp. 76-85).

Schoenmüller, V., Netzer, O., & Stahl, F. (2019). The extreme distribution of online reviews: Prevalence, drivers and implications. *Columbia Business School Research Paper*, (18-10).

Shaffer, A. (2013). Five Stars for Five Dollars: Buying Reviews, Reviewed. *The Huffington Post*.

Shang, L., Lu, Z., & Li, H. (2015). Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.

Siering, M., Koch, J. A., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems*, 33(2), 421-455.

Siponen, M. (2004). A justification for software rights. *Acm Sigcas Computers and Society*, 34(3), 3-3.

Siponen, M., Vance, A., & Willison, R. (2012). New insights into the problem of software piracy: The effects of neutralization, shame, and moral beliefs. *Information & Management*, 49(7-8), 334-341.

Slee, T. (2014). In praise of fake reviews.

Smith, H. J. (2002). Ethics and information systems: Resolving the quandaries. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 33(3), 8-22.

Smithers, R. (2011). Terms and conditions: not reading the small print can mean big problems. *The Guardian*, 11, 05-11.

Soliman, W. (2015). People-Driven, ICT-Enabled Innovation: Crowdsourcing.

Song, J., Lee, S., & Kim, J. (2015, October). Crowdtarget: Target-based detection of crowdturfing in online social networks. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 793-804).

Song, J., Lee, S., & Kim, J. (2015, October). Crowdtarget: Target-based detection of crowdturfing in online social networks. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security* (pp. 793-804).

Straub Jr, D. W., & Nance, W. D. (1990). Discovering and disciplining computer abuse in organizations: A field study. *MIS quarterly*, 45-60.

Thai, M. T., Wu, W., & Xiong, H. (Eds.). (2016). *Big data in complex and social networks*. CRC Press.

Tian, R. Y., Zhang, X. F., & Liu, Y. J. (2015). SSIC model: A multi-layer model for intervention of online rumors spreading. *Physica A: Statistical Mechanics and its Applications*, 427, 181-191.

Wang, G., Wang, T., Zheng, H., & Zhao, B. Y. (2014). Man vs. machine: Practical adversarial detection of malicious crowdsourcing workers. In *23rd USENIX Security Symposium (USENIX Security 14)* (pp. 239-254).

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2012, April). Serf and turf: crowdturfing for fun and profit. In *Proceedings of the 21st international conference on World Wide Web* (pp. 679-688).

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2012, April). Serf and turf: crowdturfing for fun and profit. In *Proceedings of the 21st international conference on World Wide Web* (pp. 679-688).

Wang, J. N., Du, J., & Chiu, Y. L. (2020). Can online user reviews be more helpful? Evaluating and improving ranking approaches. *Information & Management*, 57(8), 103281.

Wang, T., Wang, G., Li, X., Zheng, H., & Zhao, B. Y. (2013, August). Characterizing and detecting malicious crowdsourcing. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM* (pp. 537-538).

Willison, R., & Warkentin, M. (2013). Beyond deterrence: An expanded view of employee computer abuse. *MIS quarterly*, 1-20.

Wu, L., & Liu, H. (2018). Detecting Crowdturfing in Social Media.

Yang, X., Yang, Q., & Wilson, C. (2015). Penny for your thoughts: Searching for the 50 cent party on sina weibo. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 9, No. 1, pp. 694-697).

Διπλωματική Εργασία: Crowdturfing και E-commerce: Επισκόπηση και μελέτη των επιπτώσεων σε συστήματα ηλεκτρονικού εμπορίου

Yao, Y., Viswanath, B., Cryan, J., Zheng, H., & Zhao, B. Y. (2017, October). Automated crowdturfing attacks and defenses in online review systems. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security* (pp. 1143-1158).

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: an introduction*. Cambridge University Press.