Αναφορά Trusted Flagger για το έτος 2025

Η παρούσα αναφορά αποτυπώνει τη δραστηριότητα του Greece Fact Check κατά το έτος 2025 στο πλαίσιο της ιδιότητάς του ως Trusted Flagger βάσει του Κανονισμού για τις Ψηφιακές Υπηρεσίες (Digital Services Act), με αποκλειστική γεωγραφική εστίαση στην Ελλάδα. Το Greece Fact Check αποτελεί πλήρως ανεξάρτητο οργανισμό και ανήκει στην αστική μη κερδοσκοπικής εταιρείας “Παρατηρητήριο κατά της Παραπληροφόρησης”. Δεν συνδέεται με κανένα τρόπο χρηματοδότησης ή λειτουργίας με οποιαδήποτε επιγραμμική πλατφόρμα.

Με την υπ’ αριθμ. πρωτ. ΕΕΤΤ 1134/18 απόφαση της 25ης Σεπτεμβρίου 2024 της Εθνικής Επιτροπής Τηλεπικοινωνιών και Ταχυδρομείων, αποδόθηκε στο Greece Fact Check η ιδιότητα της αξιόπιστης πηγής επισήμανσης παράνομου περιεχομένου (Trusted Flagger), σύμφωνα με το άρθρο 22 του Κανονισμού (ΕΕ) 2022/2065. Η αναγνώριση αυτή βασίστηκε στην εμπειρία, την ανεξαρτησία και την αποδεδειγμένη ικανότητα του οργανισμού να εντοπίζει και να τεκμηριώνει επιβλαβές περιεχόμενο που επηρεάζει τους χρήστες στην Ελλάδα.

Κατόπιν της απόφασης αυτής, το Greece Fact Check αναγνωρίστηκε επισήμως ως Trusted Flagger από σειρά πολύ μεγάλων επιγραμμικών πλατφορμών. Κατά το πρώτο εξάμηνο του 2025, οι πλατφόρμες αυτές δημιούργησαν ειδικά κανάλια επικοινωνίας και παρείχαν γενικές οδηγίες για τη διαδικασία υποβολής αναφορών.

Ειδικότερα, η αναγνώριση πραγματοποιήθηκε από τη Google στις 24 Φεβρουαρίου 2025, τη Dailymotion την 1η Μαρτίου 2025 και τη Meta στις 19 Μαρτίου 2025, ενώ ακολούθησαν οι AliExpress, X, eBay, TikTok και Snapchat έως τα τέλη Απριλίου 2025. Οι αναγνωρίσεις αυτές κατέστησαν δυνατή την ουσιαστική ενεργοποίηση του ρόλου του Trusted Flagger κατά τη διάρκεια του έτους.

Meta

Από τις 24 Μαρτίου έως τις 31 Δεκεμβρίου 2025, το μεγαλύτερο μέρος της δραστηριότητας του Greece Fact Check ως Trusted Flagger επικεντρώθηκε στις πλατφόρμες της Meta, όπου παρατηρήθηκε ιδιαίτερα αυξημένος όγκος απατηλού και επιβλαβούς περιεχομένου. Κατά το διάστημα αυτό υποβλήθηκαν συνολικά 778 αναφορές προς τη Meta. Πολλές από τις αναφορές περιλάμβαναν περισσότερες από μία περιπτώσεις διαφορετικού περιεχομένου, κυρίως από το Facebook και σε μικρότερο βαθμό από το Instagram. Συνολικά, οι αναφορές αφορούσαν 4.177 μοναδικούς συνδέσμους, οι οποίοι στην πλειονότητά τους ήταν χορηγούμενο περιεχόμενο, δηλαδή διαφημίσεις. Το περιεχόμενο αυτό αφορούσε σχεδόν αποκλειστικά απάτες οικονομικού και ιατρικού χαρακτήρα.

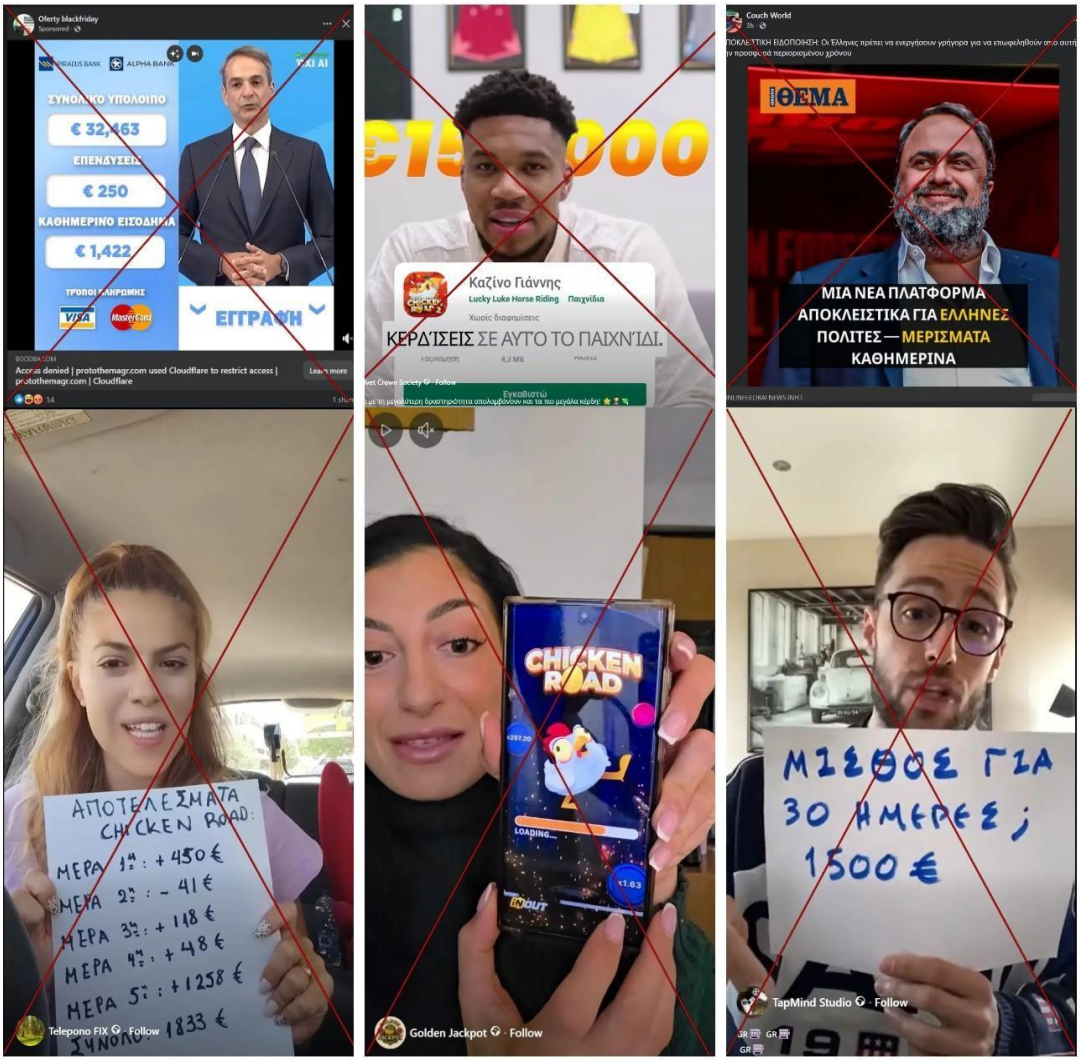
Πέραν των μεμονωμένων περιστατικών απάτης, το Greece Fact Check υπέβαλε επιπλέον 2.444 μοναδικούς συνδέσμους από τις βιβλιοθήκες διαφημίσεων των σελίδων που προωθούσαν το συγκεκριμένο περιεχόμενο. Έτσι, ο συνολικός αριθμός των συνδέσμων που υποβλήθηκαν στη Meta ανήλθε σε 6.621. Στις περισσότερες περιπτώσεις, οι εμπλεκόμενες σελίδες προέβαλαν πολλαπλές διαφημίσεις, συχνά σε μεγάλη κλίμακα και για παρατεταμένο χρονικό διάστημα. Με τον τρόπο αυτό, το Greece Fact Check ζητούσε από τη Meta να μην εξετάζει μόνο μεμονωμένες διαφημίσεις, αλλά και τις ίδιες τις σελίδες και τη διαφημιστική τους δραστηριότητα συνολικά.

Η πρακτική αυτή βασίστηκε στην εμπειρία ότι η αφαίρεση μεμονωμένων απατηλών διαφημίσεων με παράνομο περιεχόμενο, δεν οδηγούσε πάντοτε σε ουσιαστική αντιμετώπιση του προβλήματος. Με τη συστηματική αναφορά των βιβλιοθηκών διαφημίσεων, κατέστη δυνατό να ελεγχθούν και, όπου κρίθηκε απαραίτητο, να αφαιρεθούν σελίδες που διαχειρίζονταν δεκάδες, εκατοντάδες ή ακόμη και χιλιάδες απατηλές διαφημίσεις, αυξάνοντας σημαντικά τον αντίκτυπο της παρέμβασης.

Σύμφωνα με τις απαντήσεις της Meta, 697 από τις 778 αναφορές που υποβλήθηκαν οδήγησαν σε αφαίρεση περιεχομένου, ποσοστό 89,58%, ενώ 81 αναφορές, ήτοι 10,41%, δεν είχαν ως αποτέλεσμα την αφαίρεση. Τα στοιχεία αυτά καταδεικνύουν την

αποτελεσματικότητα της στοχευμένης και τεκμηριωμένης αναφοράς σε περιπτώσεις συστηματικής κατάχρησης.

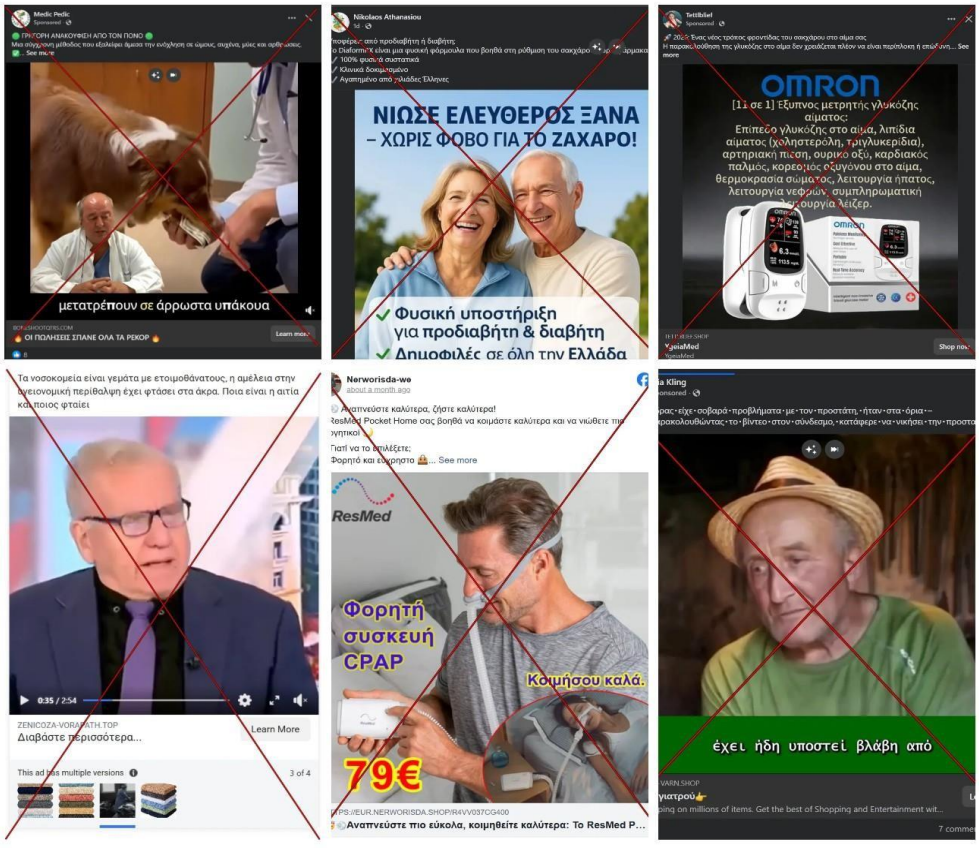
Παράλληλα, η μη αφαίρεση περιεχομένου δεν συνεπάγεται κατ' ανάγκη εσφαλμένη αξιολόγηση εκ μέρους της πλατφόρμας. Σε ορισμένες περιπτώσεις ενδέχεται να υφίσταται διαφορετική εκτίμηση ως προς το αν η κάθε περίπτωση συνιστά παράνομο ή επιβλαβές περιεχόμενο, μεταξύ του υποβάλλοντος την αναφορά και του αρμόδιου οργάνου που την εξετάζει. Η διαφοροποίηση αυτή μπορεί να οφείλεται σε διαφορετική στάθμιση των πραγματικών περιστατικών ή στην αξιολόγηση των διαθέσιμων στοιχείων και δεν αναιρεί τον ρόλο της τεκμηριωμένης επισήμανσης στο πλαίσιο του μηχανισμού Trusted Flagger.



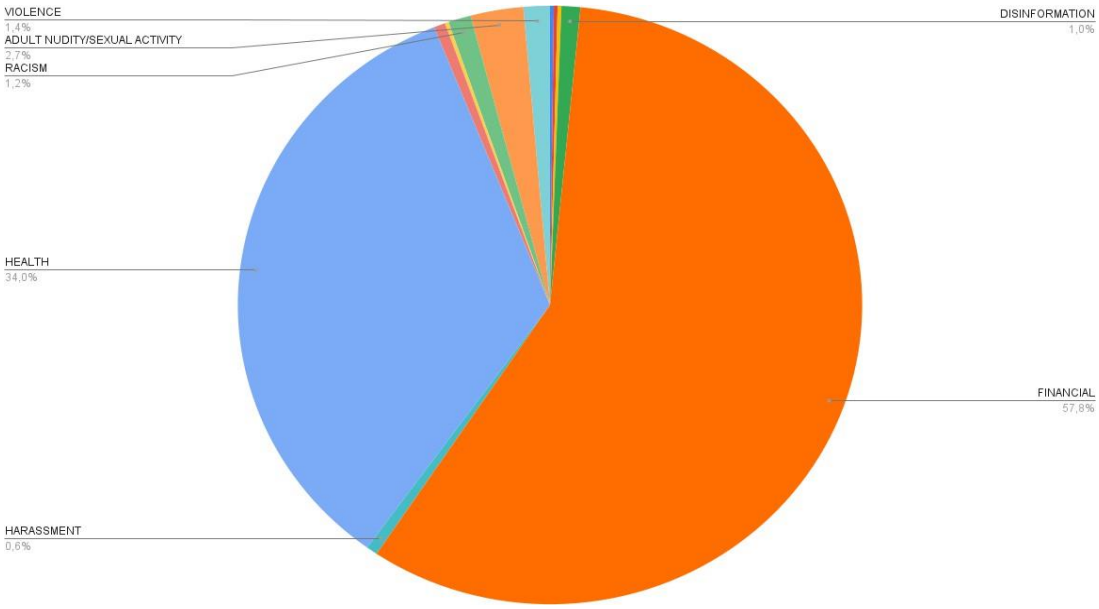
Οι οικονομικές απάτες αποτέλεσαν τη συχνότερη κατηγορία περιεχομένου που εντοπίστηκε. Κυρίαρχο ρόλο είχαν deepfake βίντεο και παραποιημένο υλικό γνωστών

προσώπων, όπως πολιτικών, κυβερνητικών στελεχών, τραπεζιτών, επενδυτών και διαφόρων διασημοτήτων, που χρησιμοποιούνταν για την προώθηση απατηλών επενδυτικών σχημάτων. Συχνά γίνεται χρήση ονομάτων και σημάτων γνωστών, mainstream ΜΜΕ, για παράδειγμα του “Πρώτου Θέματος”, επιχειρήσεων, όπως τα σούπερ μάρκετ “Σκλαβενίτης”, τραπεζών, όπως η Εθνική Τράπεζα, η Πειραιώς, η Alpha Bank και οργανισμών όπως ο ΟΠΑΠ, σε περιπτώσεις παράνομου στοιχηματισμού και τζόγου. Παράλληλα, καταγράφηκε εκτεταμένη χρήση περιεχομένου που είχε δημιουργηθεί με τεχνητή νοημοσύνη, καθώς και επαναλαμβανόμενα βίντεο ατόμων που ισχυρίζονταν ότι κέρδισαν μεγάλα χρηματικά ποσά μέσω εφαρμογών τυχερών παιχνιδιών.

Οι ιατρικές απάτες αφορούσαν κυρίως προϊόντα που διαφημίζονται παρανόμως ως θεραπείες για χρόνιες παθήσεις ή ως «ιατρικές» συσκευές χωρίς επιστημονική τεκμηρίωση. Συχνά χρησιμοποιούνταν deepfakes ή υλικό τεχνητής νοημοσύνης με τη μορφή γνωστών Ελλήνων γιατρών, ή άλλων διάσημων προσωπικοτήτων, προκειμένου να ενισχυθεί η αξιοπιστία των ψευδών ισχυρισμών.



Σε πολύ μικρότερο βαθμό, εντοπίστηκαν και αναφέρθηκαν περιπτώσεις πορνογραφικού περιεχομένου, ρατσισμού, παρενόχλησης και βίας. Παράλληλα, καταγράφηκαν περιστατικά παραπληροφόρησης τα οποία αναφέρθηκαν, αλλά δεν οδήγησαν σε ενέργειες εκ μέρους της Meta. Τα περιστατικά αυτά, συμπεριλαμβανομένων περιπτώσεων κρατικά υποκινούμενης προπαγάνδας,παρουσίαζαν ιδιαίτερα υψηλή συχνότητα και επαναληπτικότητα κατά την εξεταζόμενη περίοδο. Ωστόσο, η υποβολή σχετικών αναφορών από την πλευρά μας σταδιακά περιορίστηκε, καθώς δεν διαπιστώθηκε ουσιαστική ανταπόκριση ή συστηματική λήψη μέτρων από την πλατφόρμα κατόπιν αυτών των αναφορών. Στην πλειονότητα των περιπτώσεων, τα περιστατικά παραπληροφόρησης ελέγχθηκαν από fact- checkers σε μεταγενέστερο στάδιο, στα πλαίσια του προγράμματος 3PFC, μέσω συστηματικής παρακολούθησης και ανάλυσης του περιεχομένου που κυκλοφορούσε στην πλατφόρμα.



Γεωγραφική προέλευση απατών

Παρότι κατά τη διάρκεια της περιόδου αναφοράς δεν πραγματοποιήθηκε συστηματική καταγραφή των χωρών προέλευσης των απατών, η συνολική εμπειρία από την παρακολούθηση και ανάλυση των περιστατικών επιτρέπει την ασφαλή διαπίστωση ότι η συντριπτική πλειονότητα των διαχειριστών των εμπλεκόμενων σελίδων και λογαριασμών δεν εδρεύει σε χώρες της Ευρωπαϊκής Ένωσης. Η εικόνα αυτή προέκυψε από

επαναλαμβανόμενα μοτίβα, στοιχεία λογαριασμών και ενδείξεις που συνοδεύουν τη διαφημιστική δραστηριότητα.

Υφαρπαγή σελίδων

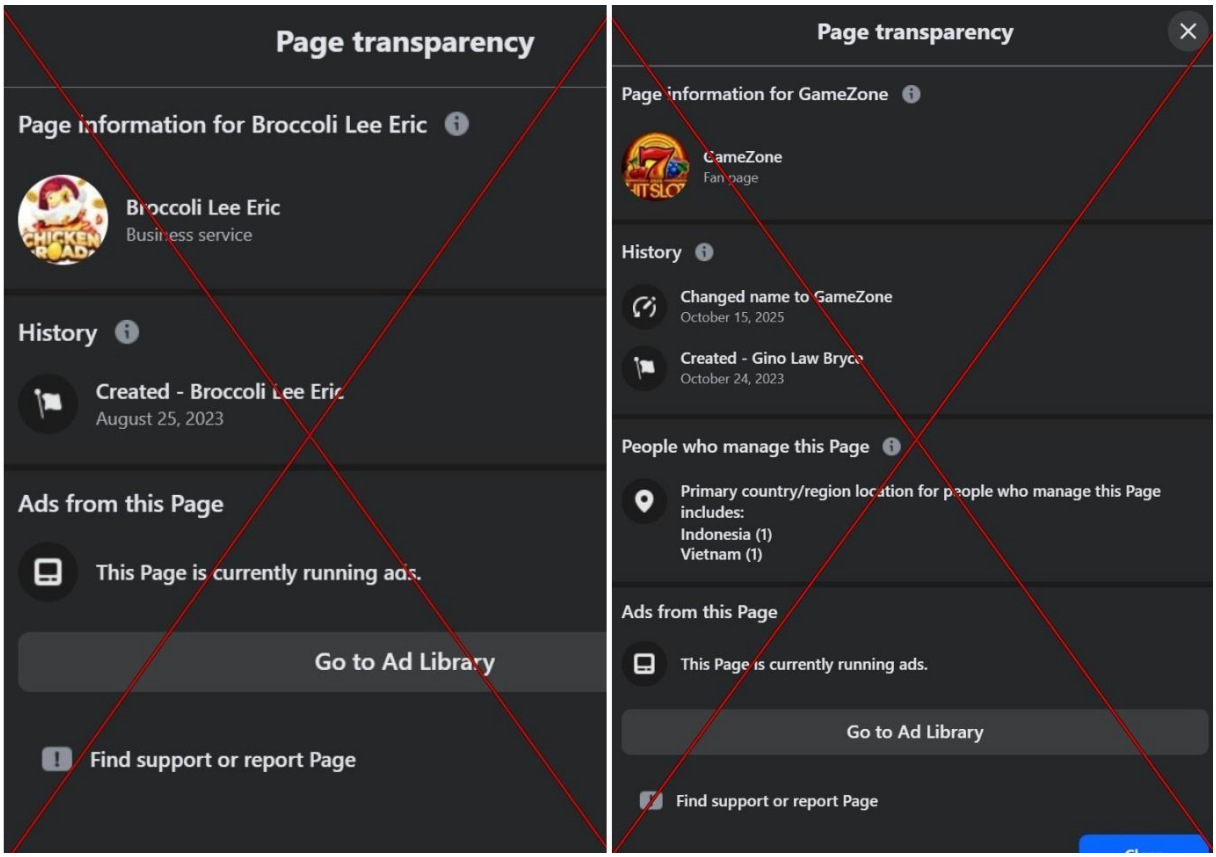
Κατά την περίοδο αναφοράς, διαπιστώθηκε επανειλημμένα το φαινόμενο υφαρπαγής/υποκλοπής σελίδων στο Facebook, μέσω απόκτησης μη εξουσιοδοτημένης πρόσβασης στα στοιχεία διαχείρισής τους. Οι σελίδες αυτές, αφού περιέρχονταν στον έλεγχο τρίτων, μετονομάζονταν και επαναπροσδιορίζονταν ως προς το περιεχόμενό τους, προκειμένου να χρησιμοποιηθούν στη συνέχεια για τη διάδοση απατηλών διαφημίσεων.

Η πρακτική αυτή επιτρέπει στους δράστες να αξιοποιούν ήδη υπάρχουσες σελίδες με ιστορικό δραστηριότητας, ακολούθους και φαινομενική αξιοπιστία, παρακάμπτοντας σε σημαντικό βαθμό τους θεωρητικούς αρχικούς μηχανισμούς ελέγχου που υποτίθεται ότι εφαρμόζονται σε νεοσύστατες σελίδες. Ωστόσο, το τελευταίο φαίνεται ότι πρακτικά δεν ισχύει.

Το φαινόμενο της υφαρπαγής σελίδων ενισχύει τη συστημική διάσταση του κινδύνου που σχετίζεται με την απατηλή διαφήμιση και αναδεικνύει την ανάγκη για αυστηρότερους ελέγχους σε περιπτώσεις αιφνίδιων αλλαγών ιδιοκτησίας, ονομασίας ή δραστηριότητας σελίδων, ιδίως όταν αυτές συνδυάζονται με μαζική διαφημιστική δραστηριότητα.

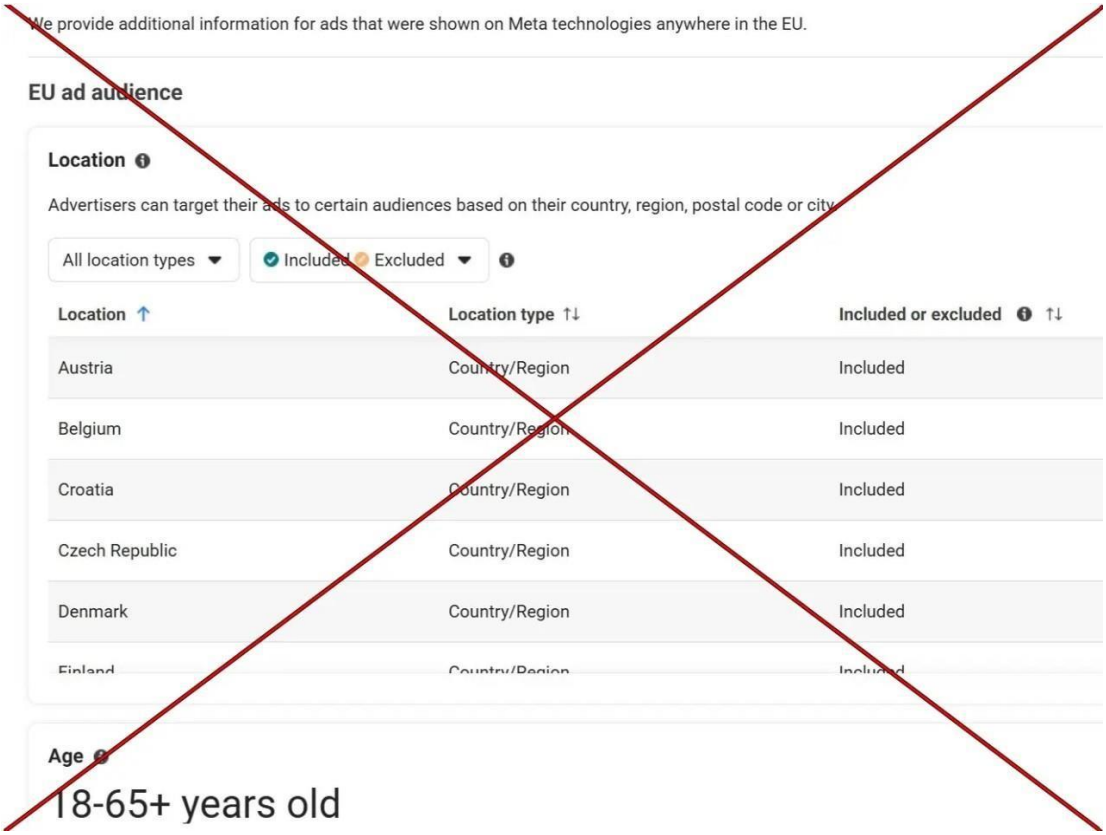
Σε αρκετές περιπτώσεις, ωστόσο, η Meta δεν παρείχε πληροφορίες σχετικά με τη χώρα προέλευσης των διαχειριστών, γεγονός που περιορίζει τη δυνατότητα ακριβέστερης γεωγραφικής αποτύπωσης. Η έλλειψη αυτή δυσχεραίνει την πλήρη κατανόηση της διασυνοριακής διάστασης των απατηλών δικτύων και υπογραμμίζει τη σημασία μεγαλύτερης διαφάνειας ως προς τα βασικά χαρακτηριστικά των οντοτήτων που αξιοποιούν τα διαφημιστικά συστήματα της πλατφόρμας.

Ακολουθεί παράδειγμα σελίδας όπου δεν φαίνεται από πού είναι οι διαχειριστές και παράδειγμα σελίδας που εκλάπη για να χρησιμοποιηθεί για απάτες.



Γεωγραφική διασπορά

Τα διαφημιστικά εργαλεία της Meta παρέχουν στους διαφημιζόμενους τη δυνατότητα να προσαρμόζουν το περιεχόμενο και τη στόχευσή τους με μεγάλη ακρίβεια, ανάλογα με τους στόχους που επιδιώκουν. Μια διαφήμιση μπορεί να εμφανίζεται σε πολλαπλές εκδοχές, ενώ μπορεί να παραμετροποιείται περαιτέρω ως προς το ηλικιακό εύρος των χρηστών αλλά και ως προς τη γεωγραφική κατανομή του κοινού στο οποίο απευθύνεται. Η ευελιξία αυτή, αν και θεμιτή για νόμιμες εμπορικές χρήσεις, αξιοποιείται συχνά και από κακόβουλους φορείς.

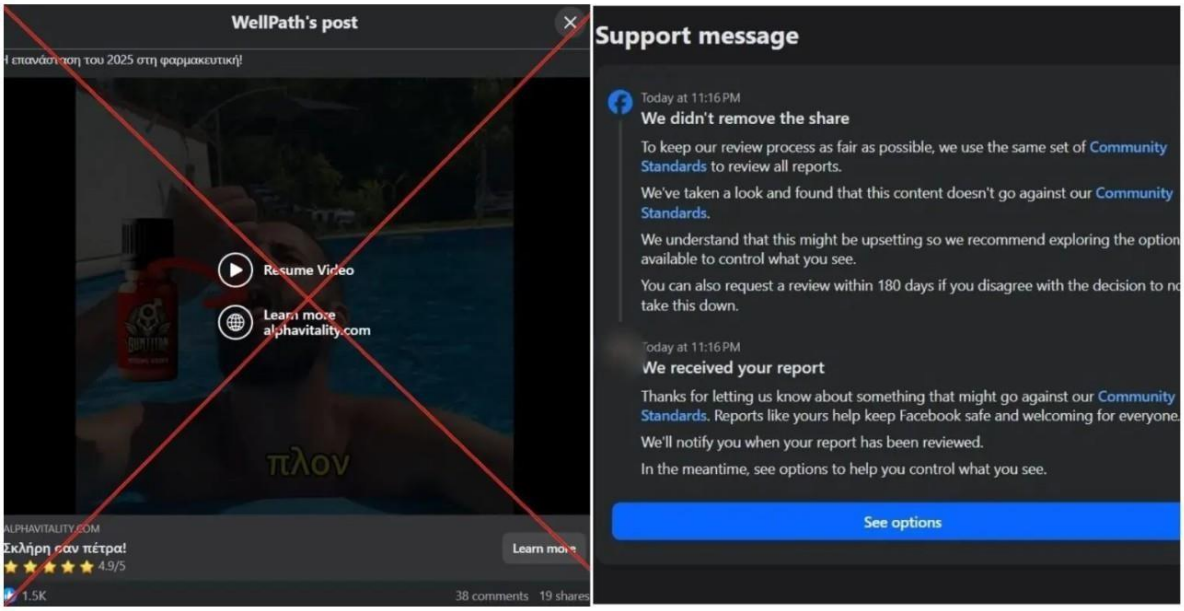


Κατά τη διάρκεια της περιόδου αναφοράς, παρατηρήθηκε επανειλημμένα το φαινόμενο κακόβουλων σελίδων να στοχεύουν ταυτόχρονα χρήστες σε περισσότερες από μία χώρες με το ίδιο ή παρόμοιο διαφημιστικό περιεχόμενο. Σε αρκετές περιπτώσεις, το ίδιο απατηλό υλικό εντοπίστηκε να προωθείται παράλληλα σε πολλαπλά κράτη-μέλη της Ευρωπαϊκής Ένωσης, συμπεριλαμβανομένης της Ελλάδας. Το περιεχόμενο αυτό, δηλαδή οι σελίδες, αναφέρθηκαν και, σε αρκετές περιπτώσεις, αφαιρέθηκαν κατόπιν ελέγχου από την πλατφόρμα.

Η πρακτική αυτή καταδεικνύει ότι οι απατηλές διαφημιστικές καμπάνιες δεν περιορίζονται εντός εθνικών συνόρων, αλλά συχνά λειτουργούν με διασυνοριακή λογική, αξιοποιώντας τα εργαλεία γεωγραφικής στόχευσης για τη μεγιστοποίηση της εμβέλειας και του οικονομικού οφέλους. Η γεωγραφική αυτή διασπορά ενισχύει τη συστημική διάσταση του κινδύνου και αναδεικνύει την ανάγκη συντονισμένης αντιμετώπισης, τόσο σε επίπεδο πλατφορμών όσο και σε επίπεδο εποπτικών αρχών.

Αναφορές από χρήστες

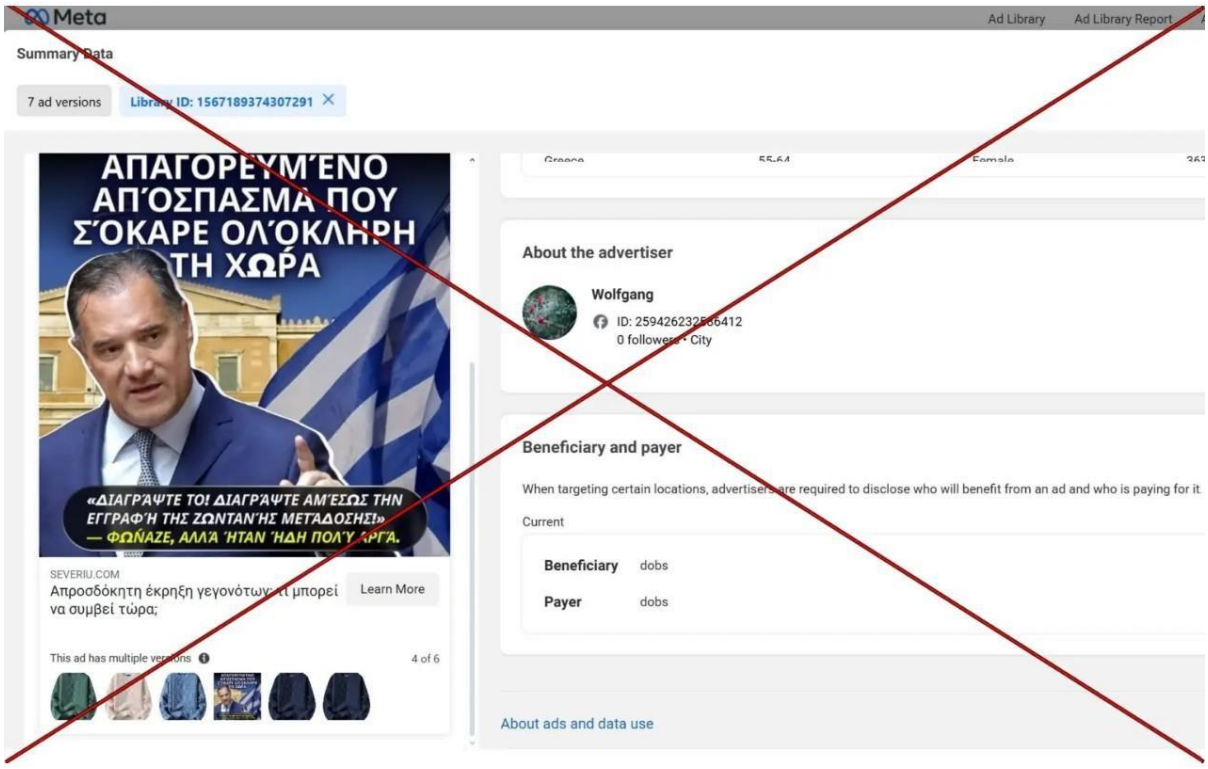
Είναι επίσης σημαντικό να επισημανθεί ότι, σε περιπτώσεις όπου το ίδιο ή παρόμοιο περιεχόμενο είχε αναφερθεί προηγουμένως στη Meta μέσω των συνήθων μηχανισμών αναφοράς που είναι διαθέσιμοι σε μεμονωμένους χρήστες, δεν διαπιστώθηκε αφαίρεσή του από την πλατφόρμα. Ακολουθεί περίπτωση πορνογραφικού περιεχομένου που υποβλήθηκε από χρήστη, ελέγχθηκε και δεν βρέθηκε να παραβιάζει τα Community Standards:



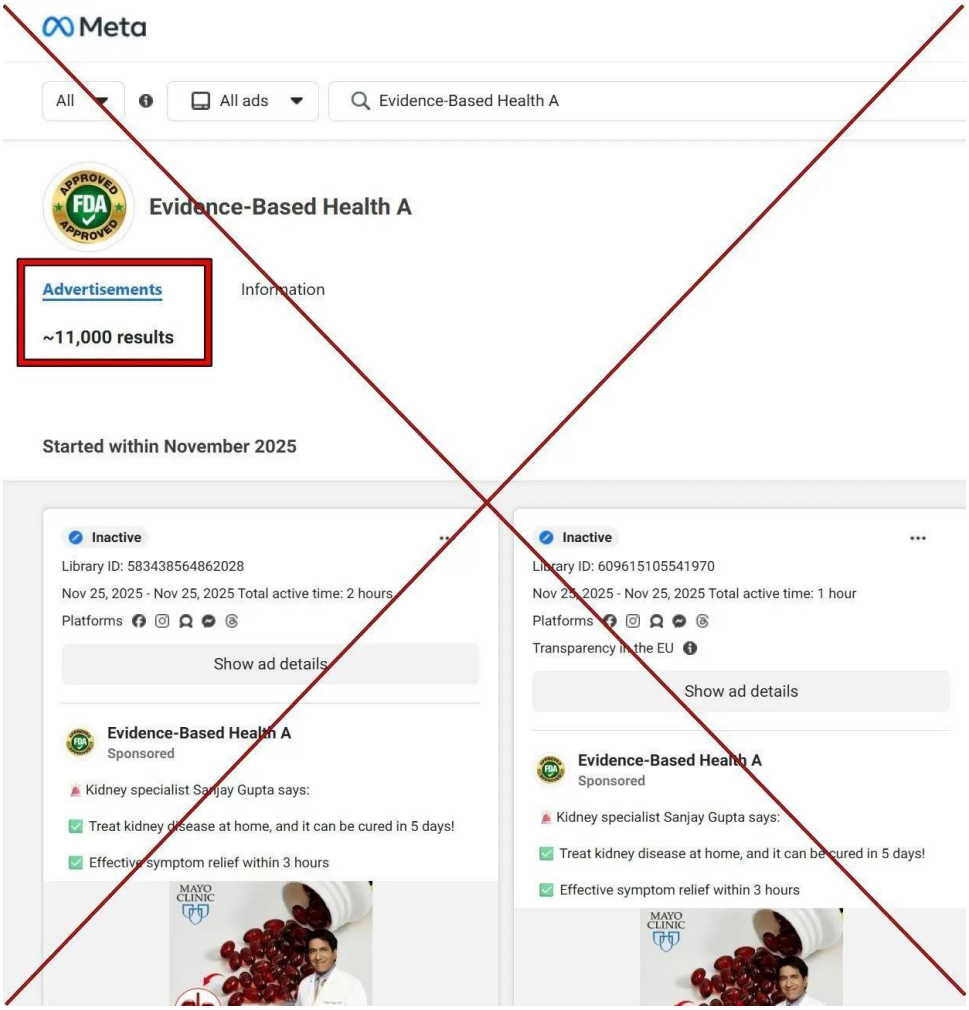
Αντιθέτως, η αφαίρεση πραγματοποιήθηκε μόνο κατόπιν υποβολής αναφοράς στο πλαίσιο της ιδιότητας ως Trusted Flaggers, διότι εκτός από προνογραφικό υλικό, περιέχει διαφήμιση για μη αδειοδοτημένο, ιατρικής φύσεως σκεύασμα. Το εύρημα αυτό αναδεικνύει αφενός την αξία του μηχανισμού Trusted Flagger, αφετέρου όμως και τους σημαντικούς περιορισμούς στην αποτελεσματικότητα των υφιστάμενων μηχανισμών διαχείρισης περιεχομένου της Meta, όταν αυτοί ενεργοποιούνται αποκλειστικά μέσω ατομικών αναφορών. Ιδίως σε περιπτώσεις χορηγούμενου περιεχομένου και οργανωμένων ή επαναλαμβανόμενων πρακτικών απάτης, η εμπειρία του Greece Fact Check υποδεικνύει ότι η έλλειψη προτεραιοποίησης και συστηματικής αξιολόγησης μπορεί να οδηγήσει σε παρατεταμένη έκθεση των χρηστών σε επιβλαβές περιεχόμενο. Από την οπτική του Digital Services Act, η παρατήρηση αυτή συνδέεται άμεσα με την ανάγκη ενίσχυσης των

μηχανισμών εκτίμησης και μετριασμού συστημικών κινδύνων εκ μέρους της Meta, ιδίως σε ό,τι αφορά τη διαχείριση διαφημιστικού περιεχομένου και τη λειτουργία των καναλιών αναφοράς.

Όσον αφορά την οικονομική διαφάνεια στις διαφημίσεις, δηλαδή ποιος πληρώνει και ποιος είναι ο οφελούμενος, η κατάσταση είναι συνήθως θολή, διότι οι πληροφορίες που συλλέγει και αποκαλύπτει η Meta, δεν προσφέρουν διαφάνεια. Στο παρακάτω παράδειγμα βλέπουμε πως ο οφελούμενος και ο πληρωτής, δεν αποκαλύπτουν την ταυτότητά τους και αναγράφονται απλά ως “dobs”.



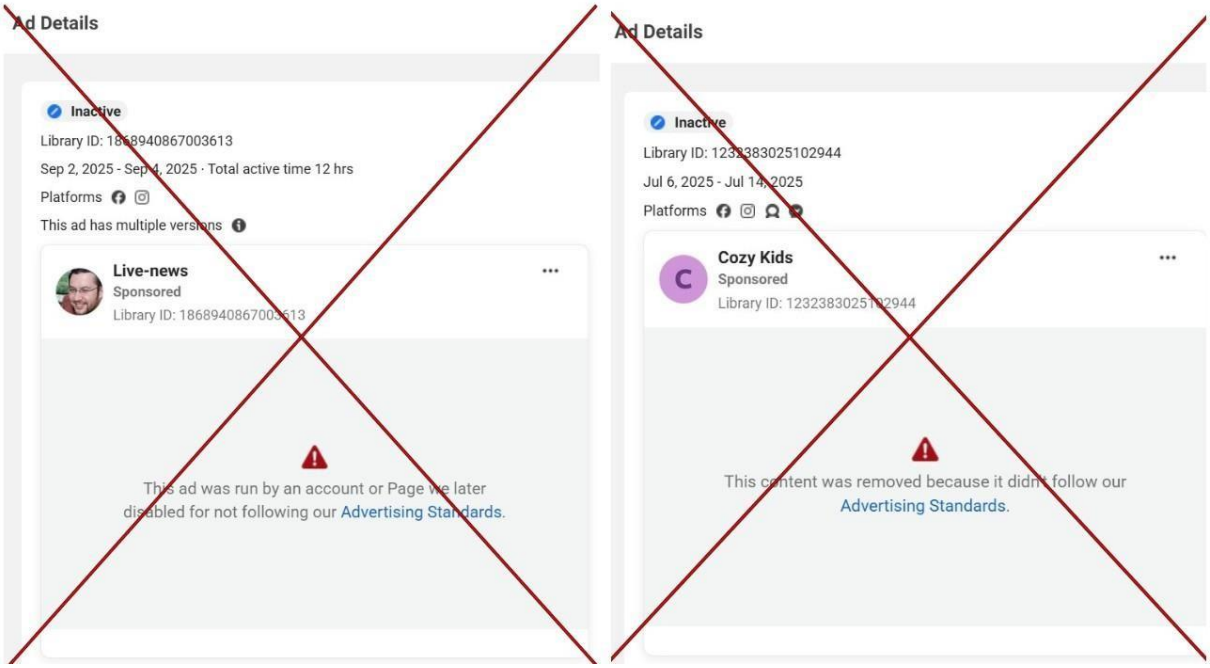
Η συνολική εικόνα που προέκυψε καταδεικνύει ότι η απατηλή διαφήμιση συνιστά συστημικό κίνδυνο στο πλαίσιο του Digital Services Act, αλλά και της εθνικής νομοθεσίας και εποπτείας. Σε πολλές περιπτώσεις σελίδες δημιουργούνταν και την ίδια ημέρα ξεκινούσαν τη μαζική προβολή απατηλών διαφημίσεων, ακόμη και δεκάδων ταυτόχρονα. Το γεγονός αυτό αναδεικνύει δομικές αδυναμίες στους μηχανισμούς ελέγχου και επιτρέπει την ταχεία εκμετάλλευση των συστημάτων διαφήμισης. Ακολουθεί παράδειγμα σελίδας που δημιούργησε 11 χιλιάδες διαφημίσεις μέσα στο διάστημα μιας εβδομάδας:



Σε ορισμένες περιπτώσεις, διαπιστώθηκε ότι η Meta είχε ήδη αφαιρέσει μεμονωμένες διαφημίσεις, με αιτιολογίες όπως ότι το περιεχόμενο δεν συμμορφωνόταν με τα Διαφημιστικά Πρότυπα της πλατφόρμας ή ότι η διαφήμιση είχε προβληθεί από λογαριασμό ή σελίδα που στη συνέχεια απενεργοποιήθηκε λόγω παραβίασης των ίδιων προτύπων. Ωστόσο, παρά τις επιμέρους αυτές ενέργειες, παρατηρήθηκε ότι οι ίδιες σελίδες μπορούσαν να συνεχίζουν να δημοσιεύουν νέες διαφημίσεις, παρόμοιου χαρακτήρα.

Το εύρημα αυτό υποδηλώνει ότι η αφαίρεση μεμονωμένων διαφημίσεων ή ακόμη και η απενεργοποίηση λογαριασμών δεν συνεπάγεται κατ' ανάγκη ουσιαστικό περιορισμό της επαναλαμβανόμενης απατηλής δραστηριότητας. Από επιχειρησιακή σκοπιά, η απουσία αυτόματης ή άμεσης σύνδεσης μεταξύ προηγούμενων παραβιάσεων και της δυνατότητας

συνέχισης της διαφημιστικής δραστηριότητας επιτρέπει την ανακύκλωση ίδιων πρακτικών και υπονομεύει την αποτελεσματικότητα των μέτρων επιβολής.



Ο ρόλος της τεχνητής νοημοσύνης (AI)

Ιδιαίτερη σημασία παρουσιάζει ο αυξανόμενος ρόλος της τεχνητής νοημοσύνης στην παραγωγή και κλιμάκωση απατηλού περιεχομένου. Κατά την εξεταζόμενη περίοδο παρατηρήθηκε συστηματική χρήση εργαλείων τεχνητής νοημοσύνης για τη δημιουργία ψευδών διαφημιστικών υλικών, πλαστών εικόνων και βίντεο, καθώς και για τη μαζική παραγωγή παραλλαγών του ίδιου αφηγήματος, με στόχο την παράκαμψη των μηχανισμών εντοπισμού των πλατφορμών. Η χρήση αυτή επιτρέπει τη γρήγορη αναπαραγωγή περιεχομένου, την προσαρμογή του σε διαφορετικά κοινά και γλώσσες και την αύξηση της φαινομενικής αξιοπιστίας των απατών, γεγονός που εντείνει τον συστημικό κίνδυνο για τους χρήστες και δυσχεραίνει την αποτελεσματική επιβολή μέτρων σε επίπεδο μεμονωμένων αναφορών. Η χρήση τεχνητής νοημοσύνης για την προώθηση παράνομου περιεχομένου ενδέχεται να αποτελεί τη σοβαρότερη απειλή για την αποτελεσματική προστασία των χρηστών στο ψηφιακό περιβάλλον. Υπό το πρίσμα αυτό, η αξιοποίηση εργαλείων τεχνητής νοημοσύνης για παράνομους σκοπούς δεν συνιστά απλώς τεχνική και τεχνολογική εξέλιξη του φαινομένου, αλλά παράγοντα που μεταβάλλει ποιοτικά και

ουσιαστικά το επίπεδο του συστημικού κινδύνου.

Σύνοψη

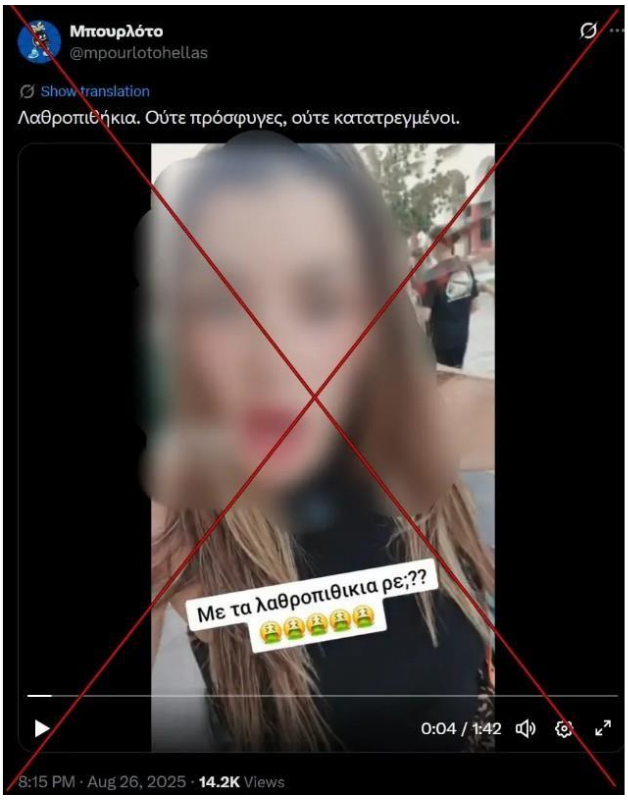
Με βάση τα ευρήματα, η απατηλή διαφήμιση στις πλατφόρμες της Meta συνιστά συστημικό κίνδυνο κατά την έννοια του Digital Services Act, καθώς δεν πρόκειται για μεμονωμένα περιστατικά αλλά για επαναλαμβανόμενες, οργανωμένες και συχνά διασυνοριακές πρακτικές που αξιοποιούν δομικά χαρακτηριστικά των διαφημιστικών και διαχειριστικών συστημάτων της πλατφόρμας. Η δυνατότητα άμεσης δημιουργίας ή υφαρπαγής σελίδων, η ταχεία έναρξη μαζικής διαφημιστικής δραστηριότητας, η στοχευμένη γεωγραφική και δημογραφική διάδοση, η περιορισμένη διαφάνεια ως προς την προέλευση των διαχειριστών και η απουσία ουσιαστικής σύνδεσης μεταξύ προηγούμενων παραβιάσεων και της δυνατότητας συνέχισης της διαφημιστικής δραστηριότητας δημιουργούν συνθήκες που ευνοούν την ανακύκλωση απατηλών πρακτικών. Παράλληλα, η χαμηλή αποτελεσματικότητα των τυπικών μηχανισμών αναφοράς από μεμονωμένους χρήστες και η αποσπασματική αφαίρεση μεμονωμένων διαφημίσεων χωρίς συστηματικό έλεγχο των εμπλεκόμενων σελίδων επιτρέπουν την παρατεταμένη έκθεση των χρηστών σε σοβαρούς, κυρίως οικονομικούς και υγειονομικούς κινδύνους. Υπό το πρίσμα του Digital Services Act, τα παραπάνω καταδεικνύουν την ανάγκη για συνεκτικότερα και προληπτικά μέτρα μετριασμού κινδύνου, τα οποία να αντιμετωπίζουν τη δομική διάσταση του φαινομένου και όχι μόνο τις επιμέρους εκδηλώσεις του.

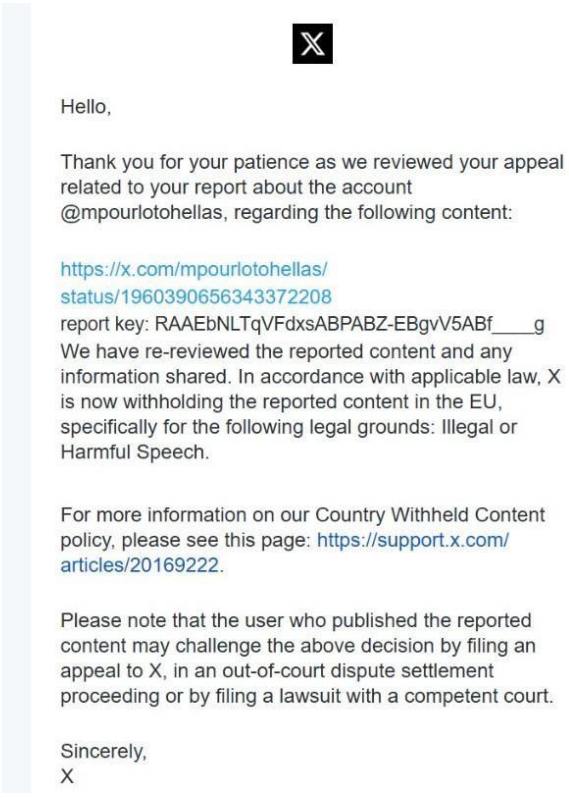
X/Twitter

Συνολικά κατατέθηκαν έντεκα αναφορές κατά την εξεταζόμενη περίοδο, οι οποίες αποτελούν τη βάση της παρούσας ανάλυσης. Οι αναφορές αυτές οδήγησαν σε διαφορετικές αποφάσεις ως προς τη λήψη μέτρων από την πλατφόρμα. Σε αρκετές περιπτώσεις κρίθηκε ότι το αναφερόμενο περιεχόμενο δεν πληρούσε τα προβλεπόμενα νομικά όρια για παρέμβαση, συμπεριλαμβανομένων πολλαπλών αποφάσεων που αφορούσαν παράνομο ή επιβλαβή περιεχόμενο, μίας περίπτωσης παραπληροφόρησης στο πλαίσιο του Νόμου για τις Ψηφιακές Υπηρεσίες και μίας περίπτωσης που εξέταζε ενδεχόμενες αρνητικές επιπτώσεις στον δημόσιο διάλογο ή στις εκλογικές διαδικασίες.

Παράλληλα, σε ορισμένες περιπτώσεις το X εφάρμοσε μέτρα ειδικά για την Ευρωπαϊκή Ένωση, προχωρώντας σε απόκρυψη του αναφερόμενου περιεχομένου εντός της ΕΕ για λόγους παράνομου ή επιβλαβούς περιεχομένου, συμπεριλαμβανομένης μίας περίπτωσης όπου η απόφαση αυτή ελήφθη κατόπιν επανεξέτασης της αναφοράς. Μία αναφορά που αφορούσε περιεχόμενο σχετικό με βία οδήγησε επίσης σε απόκρυψη του περιεχομένου εντός της ΕΕ, ενώ σε μία άλλη περίπτωση το περιεχόμενο κρίθηκε ότι παραβίαζε τους εσωτερικούς κανόνες και τις πολιτικές του X.

Ακολουθεί παράδειγμα δημοσίευσης ρατσιστικού περιεχομένου που δεν εμφανίζεται στην ΕΕ, κατόπιν αναφοράς μας. Η αρχική αναφορά μας είχε απορριφθεί, αλλά κατόπιν ένστασης το X/Twitter αποφάνθηκε το περιεχόμενο να μην εμφανίζεται σε χρήστες εντός Ευρωπαϊκής Ένωσης, με το αιτιολογικό του παράνομου ή επιβλαβούς λόγου.





Hello,

Thank you for your patience as we reviewed your appeal related to your report about the account @mpourlotohellas, regarding the following content:

https://x.com/mpourlotohellas/status/1960390656343372208
report key: RAAEbNLTqVFdxsABPABZ-EBgvV5ABf____g
We have re-reviewed the reported content and any information shared. In accordance with applicable law, X is now withholding the reported content in the EU, specifically for the following legal grounds: Illegal or Harmful Speech.

For more information on our Country Withheld Content policy, please see this page: https://support.x.com/articles/20169222.

Please note that the user who published the reported content may challenge the above decision by filing an appeal to X, in an out-of-court dispute settlement proceeding or by filing a lawsuit with a competent court.

Sincerely,
X

Με βάση τη δραστηριότητα που καταγράφηκε στην πλατφόρμα X κατά την αναφερόμενη περίοδο, διαπιστώνεται ότι, παρότι ο όγκος των αναφορών ήταν σαφώς μικρότερος σε σύγκριση με τη Meta, παραμένουν ενδείξεις συστημικού κινδύνου που σχετίζονται με την ανεπαρκή πρόληψη και αντιμετώπιση επιβλαβούς περιεχομένου. Οι περιπτώσεις που αναφέρθηκαν, αν και λίγες σε αριθμό, δεν αναιρούν το γεγονός ότι η διάδοση τέτοιου περιεχομένου μπορεί να επιτευχθεί ταχύτατα μέσω μηχανισμών αναδημοσίευσης και αλγοριθμικής ενίσχυσης. Τα στοιχεία αυτά υποδεικνύουν ότι, ακόμη και σε περιβάλλοντα με χαμηλότερο όγκο αναφορών, υφίσταται ανάγκη για ενισχυμένα μέτρα εκτίμησης και μετριασμού κινδύνου, ώστε να αποτρέπεται η ταχεία κλιμάκωση επιβλαβούς περιεχομένου και να διασφαλίζεται η αποτελεσματική προστασία των χρηστών.

Το είδος αναφοράς, ο χρόνος απόκρισης και τα ποσοστά αφαίρεσης, φαίνονται στον παρακάτω πίνακα:

Επιγραμμική πλατφόρμα / Κατηγορίες παράνομου περιεχομένου	Αριθμός ειδοποιήσεων που υποβλήθηκαν σε META	Αριθμός ειδοποιήσεων που έγιναν removed από META	Μέσος Χρόνος Απόκρισης Meta	Αριθμός ειδοποιήσεων που υποβλήθηκαν σε X	Αριθμός ειδοποιήσεων που έγιναν removed από X	Μέσος Χρόνος Απόκρισης X	Σύνολο ειδοποιήσεων που υποβλήθηκαν	Σύνολο ειδοποιήσεων που έγιναν removed
K25 Χειραγώγηση πληροφοριών και παρέμβαση από τρίτες χώρες (Foreign information manipulation and interference)	6	0	2 days - 3 months (≈26,75 days average)	1	0	1	7	1
K26 Παραποίηση πληροφοριών με στόχο να επηρεασθεί το αποτέλεσμα εκλογών (Information manipulation aimed at affecting sincerity/outcome of elections)	0	0	0	0	0	0	0	0
K29 Μη συναινετικά στοιχεία που περιέχουν βίντεο ή εικόνες με πλαστούς χαρακτήρες (deepfake) ή παρόμοια τεχνολογία που χρησιμοποιεί χαρακτηριστικά τρίτου μέρους (Nonconsensual items containing deepfake or similar technology using a third party's features)	3	1	1-35 days (≈18 days average)	0	0	0	4	1
K46 Κίνδυνος για τη δημόσια υγεία (Risk for public health)	264	235	1-5 days (≈2,4 days average)	0	0	0	264	235
K53 Ηλεκτρονικό "ψάρεμα" (Phishing)	0	0	0	0	0	0	0	0
K54 Επιχειρηματικό σχήμα πυραμίδας (Pyramid schemes)	0	0	0	0	0	0	0	0
K11 Διακρίσεις (Discrimination)	0	0	0	2	2 (1 after appeal)	2	2	2
K12 Ρητορική μίσους (Hate speech)	9	9	1-7 days (≈ 2,8 days average)	3	2	1-2	12	11
K13 Απειλές βίας (όπως απειλές θανάτου) (Threats of violence (such as death threats))	7	0	2-6 days (≈4 days average)	1	1	1 day	8	1
K14 Παραποίηση ιστορικής καταγραφής, απολογία εγκλήματος κατά της ανθρωπότητας ή άρνηση εγκλημάτων πολέμου (Historical negationism, apology of crime against humanity or war crimes denialism)	0	0	0	3	1	2 days	3	1
K30 Δημόσια παροχή πληροφοριών προσωπικής ταυτοποίησης για ένα άτομο (Doxing: publicly providing personally identifiable information about an individual)	0	0	0	0	0	0	0	0
K32 Παρακολούθηση (Stalking)	0	0	0	0	0	0	0	0
K33 Σεξουαλική παρενόχληση (sexual harassment)	0	0	0	0	0	0	0	0
K43 Πρόκληση ή υποκίνηση σε τέλεση αδικήματος επικίνδυνου για την δημόσια ασφάλεια (Provocation or incitement to commit an offense dangerous to public safety)	0	0	0	1	1	1	1	1
K72 Οργανωμένη Βία (Coordinated harm)	7	6	1-3 days (≈1,75 days average)	0	0	0	7	6
Άλλες ειδοποιήσεις παράνομου περιεχομένου στο πλαίσιο του άρθρου 16	481	450	1-10 days (2,2 average)	0	0	0	481	450

Περιορισμοί

Η άσκηση του ρόλου αυτού πραγματοποιείται εντός συγκεκριμένων θεσμικών και επιχειρησιακών περιορισμών. Ως ανεξάρτητος οργανισμός που υποβάλλει αναφορές περιεχομένου, δεν διαθέτουμε αρμοδιότητα επιβολής μέτρων ούτε πρόσβαση σε εσωτερικά δεδομένα ή συστήματα της πλατφόρμας, ενώ η τελική αξιολόγηση και απόφαση αφαίρεσής ή μη περιεχομένου, ανήκει αποκλειστικά σε αυτές. Η άσκηση του ρόλου αυτού επηρεάζεται και από περιορισμούς που σχετίζονται με τη διαθέσιμη δυναμικότητα του οργανισμού, ανεξαρτήτως πλατφόρμας. Η υποβολή αναφορών και η συστηματική

παρακολούθηση περιεχομένου σε πλατφόρμες όπως η X και η Meta πραγματοποιούνται στο πλαίσιο περιορισμένων ανθρώπινων, χρονικών και οικονομικών πόρων, γεγονός που δεν επιτρέπει την πλήρη και συνεχή κάλυψη του συνόλου της προβληματικής δραστηριότητας. Ως αποτέλεσμα, οι παρεμβάσεις βασίζονται σε διαδικασίες προτεραιοποίησης, με κριτήρια όπως η σοβαρότητα του περιεχομένου, η πιθανή του απήχηση και η συνάφειά του με ζητήματα δημόσιου ενδιαφέροντος.

Η συγκεκριμένη δραστηριότητα δεν συνοδεύεται από οικονομική αποζημίωσε και υλοποιείται μέσω της ανακατεύθυνσης ανθρώπινων, χρονικών και οικονομικών πόρων από άλλες δράσεις του οργανισμού. Ως αποτέλεσμα, η δυνατότητα συστηματικής εμπλοκής σε διαδικασίες αναφοράς και παρακολούθησης περιεχομένου σε πλατφόρμες όπως η X και η Meta εξαρτάται άμεσα από τη συνολική χρηματοδοτική βιωσιμότητα του οργανισμού. Οι συνθήκες αυτές περιορίζουν αναπόφευκτα τον όγκο και τη συχνότητα των παρεμβάσεων, χωρίς να αναιρούν τη σημασία του ρόλου αυτού για τη διαφάνεια, τη λογοδοσία και την αξιολόγηση της εφαρμογής του ενωσιακού πλαισίου και της εθνικής νομοθεσίας από τις πολύ μεγάλες διαδικτυακές πλατφόρμες.

Συμπεράσματα

Η εμπειρία του Greece Fact Check κατά το 2025, στο πλαίσιο της ιδιότητάς του ως Trusted Flagger, καταδεικνύει ότι η απατηλή διαφήμιση και το οργανωμένο επιβλαβές περιεχόμενο δεν αποτελούν μεμονωμένα περιστατικά, αλλά φαινόμενα με σαφή συστημική διάσταση. Τα ευρήματα αναδεικνύουν δομικές αδυναμίες στα συστήματα διαφήμισης και διαχείρισης περιεχομένου. Η πρακτική εμπειρία δείχνει ότι χωρίς στοχευμένες και συνεκτικές παρεμβάσεις σε επίπεδο σελίδων και διαφημιστικών υποδομών, η επιβολή παραμένει αποσπασματική και ανεπαρκής για την ουσιαστική προστασία των χρηστών.

Παράλληλα, η σημαντική διαφορά στην αποτελεσματικότητα μεταξύ των τυπικών μηχανισμών αναφοράς των χρηστών και των αναφορών Trusted Flagger υπογραμμίζει την προστιθέμενη αξία του θεσμού, αλλά και την ανάγκη ενίσχυσης των υποχρεώσεων εκτίμησης και μετριασμού συστημικών κινδύνων. Η παρούσα έκθεση καταδεικνύει ότι η ουσιαστική αντιμετώπιση του φαινομένου απαιτεί μεγαλύτερη διαφάνεια, προληπτική

προσέγγιση και αξιοποίηση εξειδικευμένης τεχνογνωσίας, ώστε οι Trusted Flaggers να λειτουργούν συμπληρωματικά και όχι αντισταθμιστικά σε αδυναμίες των μηχανισμών ελέγχου περιεχομένου (moderation).

Καθ' όλη τη διάρκεια της αναφερόμενης περιόδου, το Greece Fact Check εφάρμοσε δικλείδες ασφαλείας ώστε η άσκηση της ιδιότητας του Trusted Flagger να γίνεται με αναλογικό και υπεύθυνο τρόπο. Οι αναφορές βασίστηκαν σε εθνική νομοθεσία και Ευρωπαϊκές οδηγίες που έχουν ενσωματωθεί σε αυτη, όπως ενδεικτικά για τον παράνομο διαδικτυακό τζόγο, την κυκλοφορία μη αδειοδοτημένων φαρμακευτικών προϊόντων και συσκευών. Δεν αναφερόταν περιεχόμενο απλώς λόγω γενικής αμφιλεγόμενης φύσης, αλλά μόνο όταν υπήρχαν σαφή στοιχεία συστηματικής κατάχρησης. Επίσης. δεν σημειώθηκαν μεταβολές στην οργανωτική, διοικητική ή επιχειρησιακή λειτουργία του Greece Fact Check που να επηρεάζουν την ανεξαρτησία του κατά την άσκηση της ιδιότητάς του ως αξιόπιστης πηγής επισήμανσης παράνομου περιεχομένου. Δεν υπήρξαν αλλαγές στη νομική ή λειτουργική του δομή, στη σύνθεση ή στη σχέση του με άλλους φορείς, συμπεριλαμβανομένων επιγραμμικών πλατφορμών, ούτε στις πηγές χρηματοδότησης που θα μπορούσαν να δημιουργήσουν σύγκρουση συμφερόντων ή σχέση εξάρτησης. Ο οργανισμός συνέχισε να εφαρμόζει εσωτερικές διαδικασίες που διασφαλίζουν ότι όλες οι δραστηριότητες Trusted Flagger εκτελούνται με πλήρη λειτουργική ανεξαρτησία από τις πλατφόρμες, χωρίς παρέμβαση ή καθοδήγηση ως προς το περιεχόμενο, τον χρόνο ή το αποτέλεσμα των αναφορών, με τις αποφάσεις να λαμβάνονται αποκλειστικά βάσει εσωτερικής αξιολόγησης και τεκμηρίωσης.