

2025 Yearly report of Trusted Flagger work

March 2026



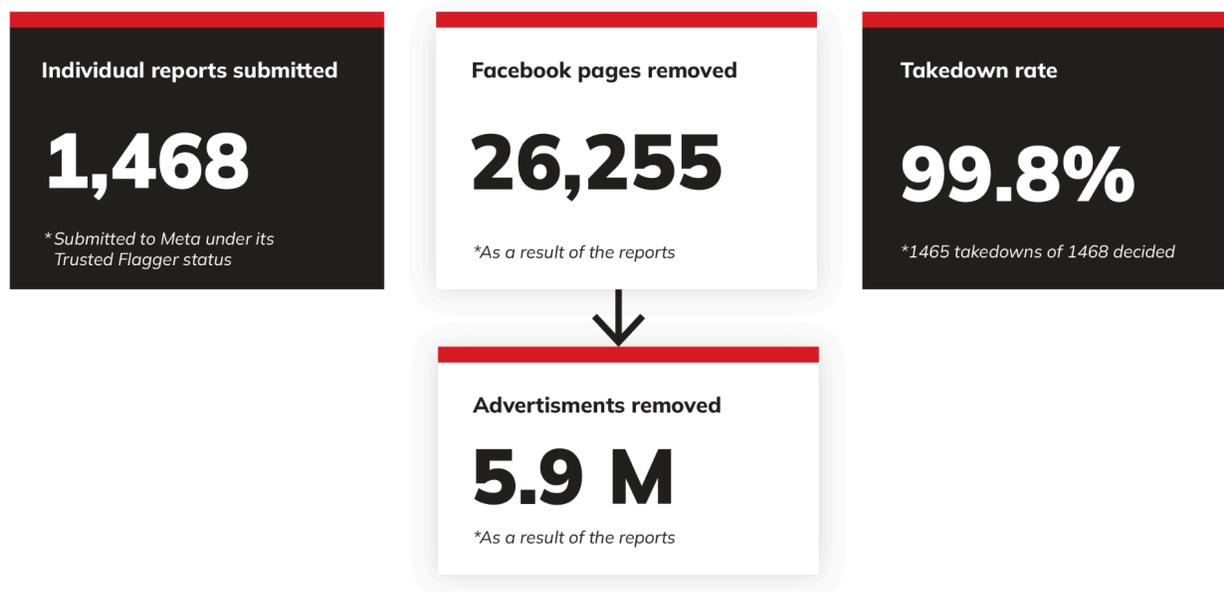
Index

- 1. Executive Summary**
- 2. Scale of the problem: A market overview**
- 3. Purpose and scope**
- 4. About Debunk.org**
- 5. Methodology**
- 6. Trusted Flagger Performance Overview and Key Statistics (2025)**
- 7. Case studies**
- 8. Conclusion**
- 9. Demonstration of Independence pursuant to Article 22(2) (b) DSA**



1. Executive Summary

Between April and December 2025, Debunk submitted **1,468 reports to Meta under its Trusted Flagger status**. These reports led to the removal of **26,255 Facebook pages** and approximately **5.9 million advertisements**. The removed content was primarily associated with **fraudulent e-commerce and investment schemes, illegal online gambling operations, and the sale of prohibited medical products**.



Meta response times exhibited significant **week-to-week** variation.

Debunk experienced systemic and recurring **operational challenges** in its use of Meta's Trusted Flagger and Official Request Portal.

Based on publicly available data from Meta's Ad Library, **5.9 million advertisements** generated an estimated aggregate EU reach of **2.34 billion** across EU Member States. In this report, 'reach' is taken directly from Meta's transparency tools.



Reach refers to the estimated number of Accounts Center accounts that saw an ad at least once, while impressions refer to the total number of times an ad was displayed. Because EU reach is reported at the ad level and then summed across millions of ads, the same individuals may be exposed to multiple ads and thus counted more than once in the aggregate figures. All reach and impression values in the following chapters should therefore be interpreted as indicative estimates rather than precise counts of unique individuals or ad views.

Debunk's reports achieved a **takedown rate of 99.8%**, demonstrating the effectiveness of structured, evidence-based reporting mechanisms. However, despite this high enforcement rate, **Meta response times exhibited significant week-to-week variation**. These findings provide robust empirical evidence of both the scale of digital harm and the operational limitations of current platform enforcement systems, offering a valuable basis for defining practical standards of responsiveness, accountability, and regulatory compliance in content moderation and platform governance frameworks.

2. Scale of the problem: A market overview

Online scams have evolved from marginal criminal activity into a global, highly sophisticated, and profitable industry. In 2024, the **total value of fraudulent payment transactions reported across the European Economic Area (EEA) reached €4.2 billion¹**, a 17% year-on-year increase.

Indeed, technology and the internet lie at the core of this ecosystem. While **online investment fraud and phishing scams targeting personal data are among the most prevalent forms of cybercrime in Europe²**, criminals increasingly benefit from advanced tools such as generative AI and algorithmic manipulation to enhance the effectiveness of their operations. These technologies enable the automation of website creation, localisation of scam content, and the deployment of deepfake audio and video to convincingly impersonate trusted individuals, significantly increasing the success rate of fraudulent schemes.

This market is extremely profitable not only for criminals but also for major digital platforms, as scammers frequently purchase advertising space to reach a broader pool of victims

¹ 2025 REPORT ON PAYMENT FRAUD, ECB, <https://www.ecb.europa.eu/press/intro/publications/pdf/ecb.ebaecb202512.en.pdf>

² Global Financial Fraud Assessment, Interpol, <https://www.interpol.int/en/Crimes/Financial-crime/Financial-crime-initiatives>



across the EU. Internal documents revealed last year showed that **Meta projected up to 10% of its 2024 revenue would derive from advertisements linked to scams and banned goods³**, while internally estimating **22 billion organic scam attempts every day**, in addition to approximately **15 billion scam advertisements served daily** (see Infographic 1).

While Debunk.org does not have access to Meta's internal financial data, these figures illustrate how platform business models can inadvertently incentivise the industrial-scale dissemination of fraudulent content.

Daily exposure on Meta *(in billions)*

December 2024



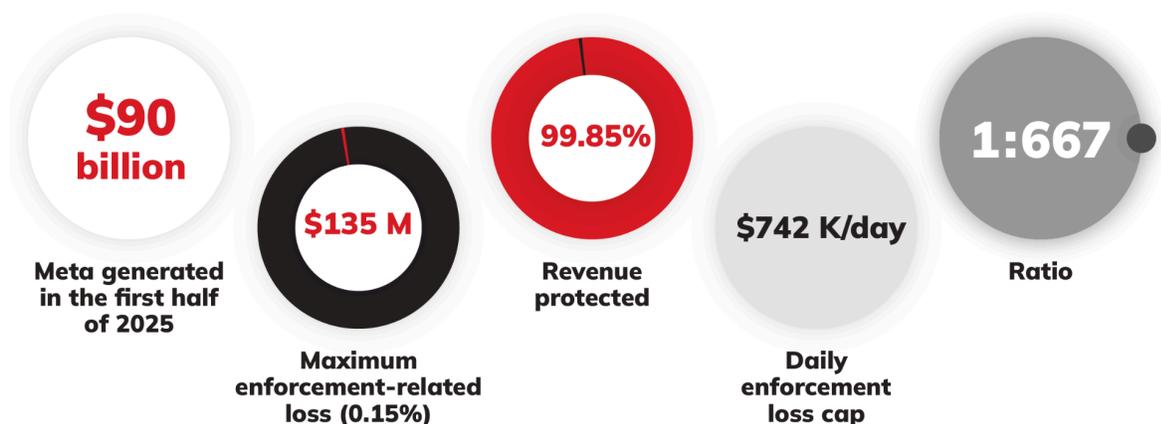
Source: Meta's internal documents leaked, according to Reuters.

Beyond the sheer scale of scam activity on its platforms, Meta's internal policies, as described in leaked documents reported by Reuters, indicate a **clear expectation of continued profitability from this ecosystem**. According to these documents, for every \$667 in revenue generated, Meta permits only \$1 to be lost as a result of enforcement actions, underscoring the exceptionally strict internal limits placed on regulatory and compliance-driven interventions.

³ Meta is earning a fortune on a deluge of fraudulent ads, documents show, Reuters, <https://www.reuters.com/investigations/meta-is-earning-fortune-deluge-fraudulent-ads-documents-show-2025-11-06/>



Infographic 2: Maximum revenue Meta allows itself to lose due to enforcement actions: \$135 million



Maximum revenue Meta allows itself to lose*: **\$135 million**

**due to enforcement actions*

Source: Meta's internal documents leaked, according to Reuters.

Following a meeting with the Chief Executive Officer, Meta executives responsible for platform integrity and enforcement established internal targets to reduce the proportion of company revenue attributable to scams, illegal gambling, and the sale of prohibited goods. According to the leaked internal documents, Meta aims to decrease this share from an estimated **10.1% in 2024 to 7.3% by the end of 2025**, followed by a further reduction to **6% by the end of 2026** and **5.8% in 2027**. While these targets signal an intention to curb reliance on illicit and harmful advertising, they also imply the continued generation of substantial revenue from these activities over the coming years.

3. Purpose and Scope

On March 23, 2025, [Debunk.org](#) was officially appointed by the Digital Services Act (DSA) supervisory authority as a trusted flagger, tasked with handling cases involving infringements of intellectual property, commercial rights and illegal online content. This annual report provides a detailed account of the organisation's activities and performance in fulfilling its responsibilities under Article 22 of the DSA.



Pursuant to Article 22(3) of the DSA, trusted flaggers are required to prepare and publish an annual report outlining the actions undertaken, the procedures applied, and the outcomes achieved in relation to the notices submitted, as well as the reports filed under Article 16 in their role as trusted flaggers.

Covering the period from 11 April to 31 December 2025, this report aims to promote transparency, accountability, and adherence to the regulatory obligations governing Debunk.org's designation as a trusted flagger.

4. About Debunk.org

Debunk.org is an independent technology-focused think tank and non-governmental organisation based in Vilnius, Lithuania, dedicated to countering online disinformation and strengthening media literacy. Founded in 2017, the organisation conducts large-scale monitoring and analysis of digital information environments, using artificial intelligence and data-driven tools to detect, track, and expose coordinated disinformation campaigns, influence operations, and online scams.

5. Methodology

Debunk.org's work in 2025 focused on advertisements published on Facebook and made publicly available through the [Meta Ads Library](#). The primary objective was to identify and document pages advertising content and goods that violate the [Meta Community Standards](#), with particular emphasis on fraud, scams, and other deceptive practices.

Using data-driven analytical tools, the organisation systematically collected and compiled these pages and advertisements. This approach enabled a comprehensive examination of both content and behavioural patterns, allowing for the identification of coordinated networks of accounts engaged in the large-scale promotion of policy-violating advertisements.

In parallel, network analysis tools were implemented to assess link-sharing behaviours across pages and advertiser profiles. By mapping connections between accounts that promoted identical fraudulent domains, Debunk.org was able to detect and document extensive, coordinated schemes. This methodological approach proved instrumental in uncovering major fraudulent networks, which are described in detail throughout the report.



To ensure methodological consistency and transparency, Debunk.org applied a standardised reporting template across all documented cases. Each case file includes a structured explanation of the findings and archived evidence of the advertisements.

Finally, Debunk.org systematically recorded the date of submission for each report filed with Meta, as well as the platform’s response time. This enabled an assessment of Meta’s responsiveness and enforcement practices, contributing to a broader evaluation of platform governance and accountability.

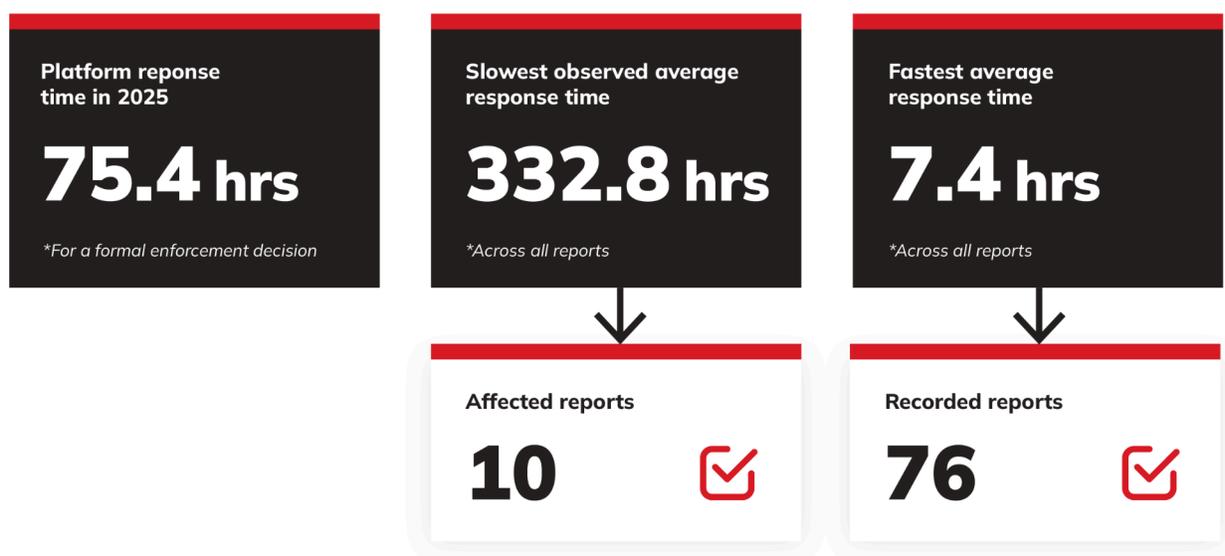
6. Trusted Flagger Performance Overview and Key Statistics (2025)

Between April and December 2025, Debunk.org submitted **a total of 1,468 reports** through Meta’s enforcement and reporting channels, covering **26,255 Facebook pages** linked to accounts engaging in fraudulent, deceptive, and illegal content.

These reporting efforts **resulted in a takedown rate of 99.8%**, with 1,465 successful enforcement actions out of 1,468 decisions, demonstrating a consistently high level of accuracy and evidentiary quality in Debunk.org’s submissions.



Despite the high takedown rate, Debunk.org’s report submissions recorded an **average platform response time of 75.4 hours** to reach a formal enforcement decision in 2025, with **significant week-to-week variation**, as illustrated in the chart. The **slowest observed average response time reached 332.8 hours**, affecting ten reports, while the **fastest average response time was 7.4 hours**, recorded across 76 reports.



Although the DSA does not establish a maximum deadline for platform action – stating only that enforcement must occur within a “reasonable time” – Debunk.org’s empirical data provides a concrete operational benchmark, creating a valuable basis for defining practical standards of responsiveness and accountability in platform enforcement.

The majority of reports submitted by Debunk.org to Meta in 2025 focused on **fraudulent e-commerce and investment schemes, illegal online gambling operations, and the sale of prohibited medical products.**

7. Case studies

- **A huge network of fake Facebook accounts pretending to be doctors: Dr X. Wonder Medicine campaign**

This operation represented the largest coordinated scam scheme reported by Debunk.org in 2025. It involved the systematic promotion of fraudulent medical products and treatments through thousands of Facebook pages, the majority of which falsely claimed to represent medical professionals, clinics, or healthcare providers, using the title “Dr.” in their username.

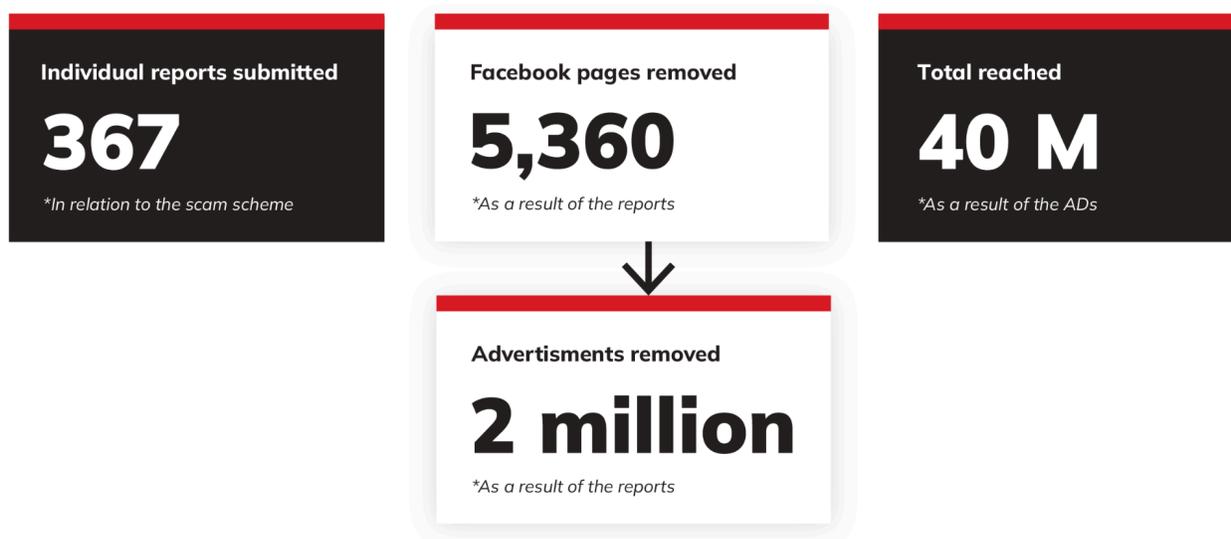
The network disseminated deceptive advertisements for unverified products targeting a wide range of health conditions, including weight loss, erectile dysfunction, ophthalmological and vision problems, orthopaedic disorders, dermatological conditions, rheumatic diseases, and anti-ageing treatments. These advertisements routinely made false and exaggerated claims, promising unrealistic outcomes such as rapid weight loss, penile enlargement, and

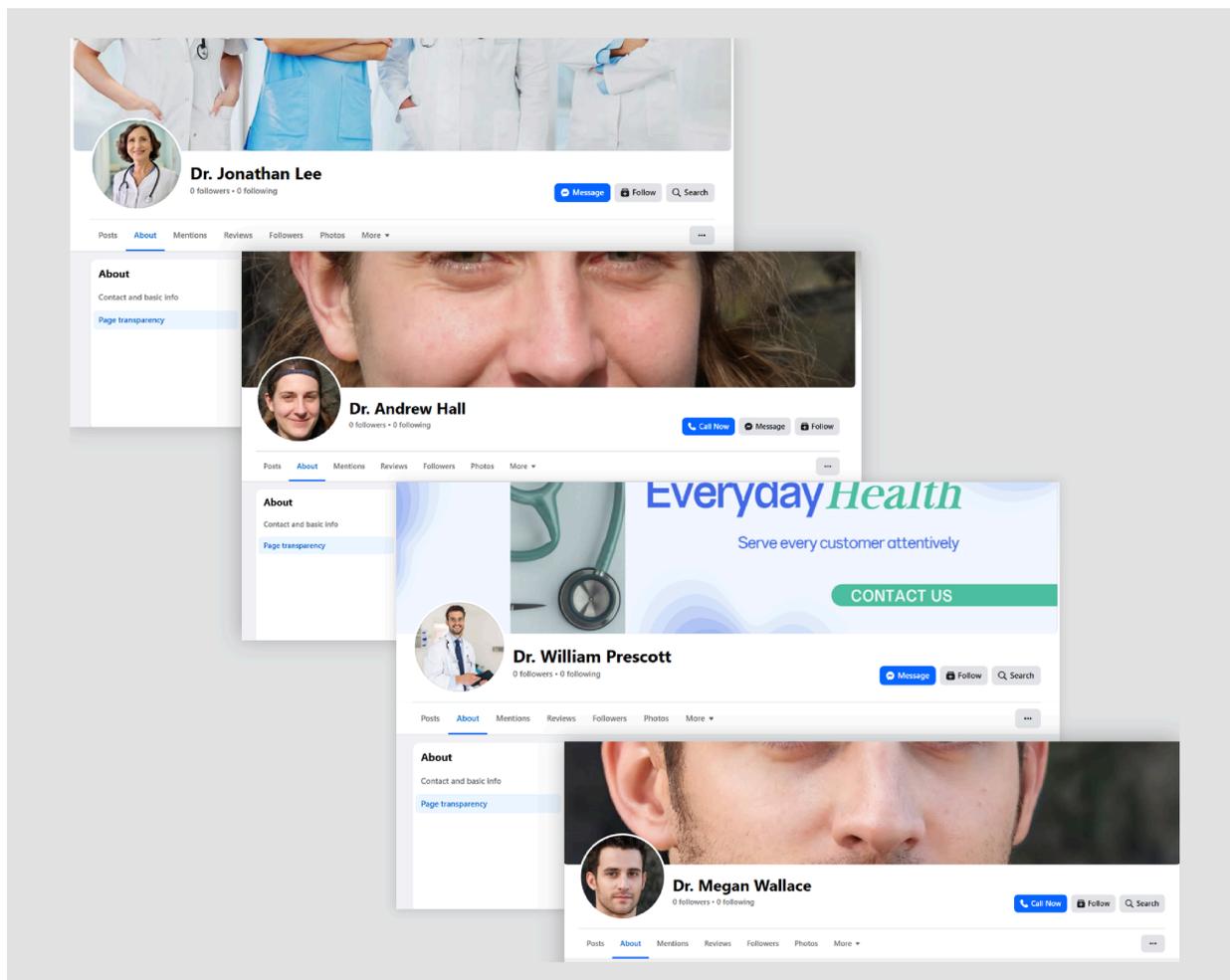


the cure or significant improvement of chronic medical conditions, without any credible evidence of safety or clinical efficacy.

Users who engaged with these advertisements were redirected to fraudulent landing pages designed to simulate legitimate news articles and medical endorsements. These pages typically featured fabricated journalistic content, forged testimonials, and manipulated images of purported medical professionals to create a false sense of credibility and authority. The content frequently exploited emotional and psychological pressure, employing fear-based messaging, social stigma, and shame-inducing language to coerce users into making purchases.

In total, Debunk.org submitted **367 individual reports** in relation to this scam scheme, resulting in the **removal of 5,360 Facebook pages and 2 million advertisements** from Meta's platforms, targeting audiences across all EU countries. **The ads reached more than 40 million people across the continent.** This figure is derived from Meta Ads Library estimates for the ads removed in connection with this scheme and should be interpreted as an indicative estimate rather than an exact count of unique individuals.





- **Impersonation of Interpol, Europol, and EU Institutions, promising “fraud refunds”**

This scam scheme represented a large-scale, highly coordinated advertising operation on Facebook and Instagram that targeted victims of previous investment and financial fraud across the EU. The campaigns are designed to exploit individuals already affected by scams, presenting false promises of compensation, asset recovery, or restitution through the impersonation of trusted law enforcement agencies, financial regulators, and institutional authorities.

The network impersonated organisations such as Interpol, Europol, the Federal Bureau of Investigation (FBI), the European Central Bank (ECB), European Union Agency for Cybersecurity (ENISA), national police forces, and financial regulators, alongside fictitious or misrepresented “law firms” and “compensation centres”. Technical and creative similarities across the ads – including shared domains, tracking links, visual templates, and branding –

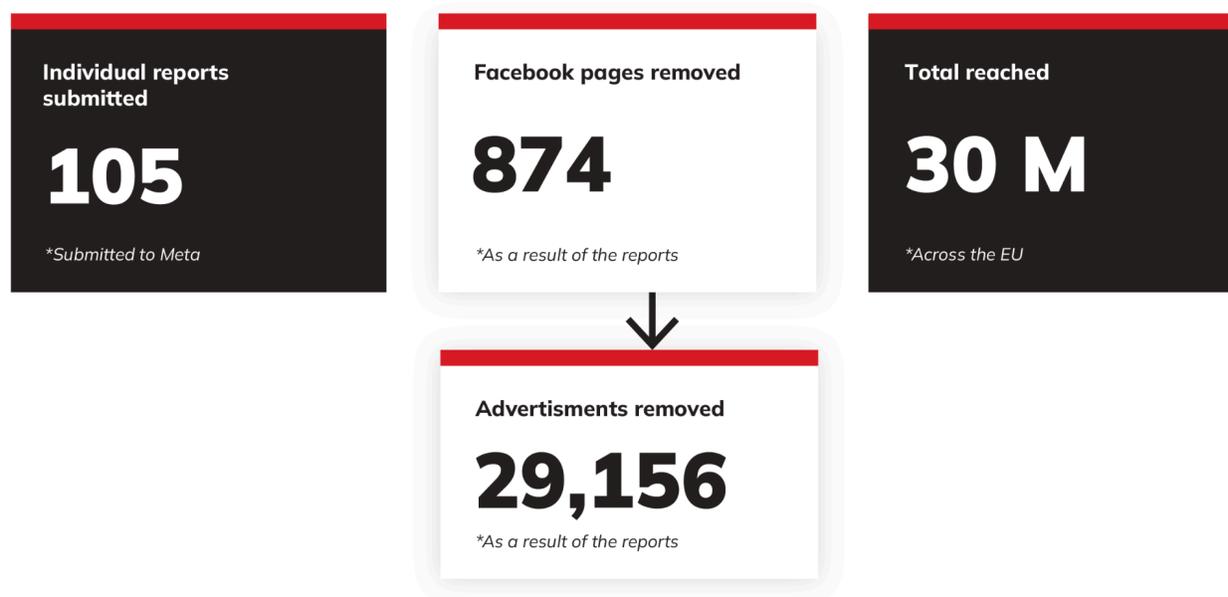


indicated the existence of a centrally managed, industrial-scale advertising infrastructure, rather than isolated or opportunistic fraudulent actors.

The ecosystem comprises more than 1,100 Facebook and Instagram pages, posting over 50,000 ads of refund and “compensation” associated with approximately 460 domains. Campaigns relied on highly targeted, short-lived micro-advertising strategies, with individual ads often remaining active for only 6 to 15 hours. This approach enabled continuous dissemination while minimising exposure to platform enforcement mechanisms.

Meta users, when clicking on the ads, were redirected to fraudulent landing pages where they were required to submit personal information. This was followed by direct contact from individuals posing as legal specialists, case managers, or cybercrime officers, who then demand upfront payments in the form of fictitious administrative fees, court costs, taxes, or verification deposits as a precondition for releasing a non-existent refund.

In response to this scheme, **Debunk.org submitted 105 detailed reports to Meta, and 874 pages were removed, along with 29,156 ads.** The ads reached an **estimated audience of 30 million users across the EU.** This figure is derived from Meta Ads Library estimates for the ads removed in connection with this scheme and should be interpreted as an indicative estimate rather than an exact count of unique individuals.





The image displays ten screenshots of Facebook advertisements for Interpol's services. The ads are arranged in two rows of five. Each ad includes the following information:

- Top Left:** Status (Active/Inactive), Library ID, and start/end dates.
- Top Right:** Platform icons (Facebook, Instagram, Messenger).
- Header:** Ad name (e.g., 'Check Investment', 'Victim Database') and 'Sponsored' label.
- Main Content:** Text in the user's language (English, Russian, or Italian) and the Interpol logo. Common text includes: "HAVE YOU BEEN SCAMMED BY A BROKER AND HAD YOUR INVESTMENT STOLEN?", "WE WILL FIND THE FRAUDSTERS AND RETURN THE STOLEN MONEY", and "APPLY FOR A REFUND".
- Bottom:** 'FB ME' and 'Learn more' or 'Apply now' buttons.

Key details from the ads include:

- Ad 1 (English):** 'Check Investment' (Library ID: 895574796468030). Text: "HAVE YOU BEEN SCAMMED BY A BROKER AND HAD YOUR INVESTMENT STOLEN?".
- Ad 2 (Russian):** 'Юридический консультант' (Library ID: 1184111663812264). Text: "Пострадали от брокера? Мы команда профессионалов с опытом работы более 5 лет. Полностью сопровождаем процедуру возврата Ваших обременений!".
- Ad 3 (Italian):** 'QuickRefund Law Firm' (Library ID: 2276873650490143). Text: "Vittima di società di investimento illegali e broker illegali? Esiste una soluzione. Ti senti frustrato dalle società di investimento e dai broker illegali?".
- Ad 4 (English):** 'Check Investment' (Library ID: 854167687155302). Text: "SPECIAL REPORT INTERNATIONAL FRAUD BROKER LIQUIDATED LIST OF VICTIMS UPDATED".
- Ad 5 (English):** 'Victim Database' (Library ID: 10346480566411). Text: "Have You Fallen Victim to Online Investment Fraud? Interpol Can Assist in Recovering Your Funds!".
- Ad 6 (Russian):** 'Check Investment' (Library ID: 1398267638685809). Text: "12 500 000\$ ИЗЪЯТЫ И БУДУТ ПЕРЕДАНЫ ПОСТРАДАВШИМ 200 БРОКЕРСКИХ КОМПАНИЙ ЗАКРЫТЫ".
- Ad 7 (Italian):** 'Fraud & Asset Recovery' (Library ID: 85874983517127). Text: "Sei stato truffato da un broker? Ti aiutiamo a recuperare il tuo denaro! Rapidamente e in modo affidabile!".
- Ad 8 (English):** 'Victim Database' (Library ID: 437389982963234). Text: "Have You Fallen Victim to Online Investment Fraud? Interpol Can Assist in Recovering Your Funds!".
- Ad 9 (Italian):** 'Fraud & Asset Recovery' (Library ID: 85874983517127). Text: "SEI STATO VITTIMA DI UNA TRUFFA FINANZIARIA ONLINE? Offriamo un servizio professionale di rimborso tramite la procedura di chargeback, con un tasso di successo del 98,8%, consulenza gratuita e risultati garantiti.". Includes 'RECUPERA I TUOI SOLDI ORA' button.
- Ad 10 (English):** 'Victim Database' (Library ID: 437389982963234). Text: "ANOTHER FRAUDULENT COMPANY SHUT DOWN". Includes "FILL THE FORM" button.



TÜCCARLAR, ANALİSTLER, BROKERLER VEYA İNTERNET DOLANDIRICILARI TARAFINDAN MALİ DOLANDIRILGIN KURBANI MI OLDUNUZ?

İnternet Sayı Şikâyet Merkezi (İCS), internet dolandırıcılığı mağdurlarına farklar içinde emin bir yolunu buldu.

İÇİS SİBER UZMANIYLA DANIŞMAK İÇİN BİR BAŞVURU GÖNDERİN

BAŞVURU GÖNDER

ORTAKLIKLAR

- Federal Soruşturma Bürosu (FBI)
- Kraliyet Kanada Atlı Polisi (RCMP)
- Uluslararası Kriminal Polis Teşkilatı (INTERPOL)
- Avrupa Polis Ofisi (Europol)

Fundusz Skonfiskowanych Środków to międzynarodowy wyspecjalizowany mechanizm finansowy, w którym gromadzone są środki, mienie i aktywa skonfiskowane brokerom i piramidom finansowym w wyniku ich działalności oszukańczej.

FCA Fund of Confiscated Asset

Rozpocznij proces odzyskiwania funduszy

Wpisz swoje imię

+48 12 345 67 89

Nieprawidłowy numer telefonu.

Adres e-mail

ROZPOCZNIJ PROCES

Fundusz jest nadzorowany przez instytucje takie jak Interpol, Europejski Bank Centralny (EBC), Financial Intelligence, Służba Skarbową i inne organy odpowiedzialne za bezpieczeństwo oraz nadzór finansowy.

Fundusz został utworzony w 2021 roku we współpracy z Interpolem i służbami specjalnymi krajów europejskich w celu ujawnienia i podlegania do odpowiedzialności organizacji oszukańczych oraz pomocy w zwrócenie środków inwestorów.

Logos: ECIPS, CYBERPOL, EUROPOLICE, I.C.I.A.

- How illegal gambling games like “Chicken Road” exploit Meta’s ad system

Throughout 2025, Debunk.org analysed a series of large-scale cases involving the promotion of illegal online gambling games across Meta’s platforms. These cases primarily



concerned casino-style games such as **Chicken Road**, **Plinko**, and **Big Bass**, which were aggressively marketed.

Relying on deceptive advertising practices, the ads often promised guaranteed monetary gains, risk-free gameplay, and immediate withdrawals, creating a false perception of safety, profitability, and legitimacy.

Beyond misleading claims, Debunk.org identified significant regulatory breaches linked to cross-border distribution. The gambling games were aggressively promoted to users across multiple European countries, including jurisdictions in which the game developers or operators did not hold valid licences to offer gambling services. This practice exposed users to unregulated platforms and increased the risks of financial harm, fraud, and exploitation.

One of the most emblematic and large-scale cases identified by Debunk.org involved the game **Chicken Road**, a gambling application deceptively presented as a harmless casual game. The ads promised “guaranteed” winnings of up to €20,000 and exaggerated returns, including claims of x101 profits, while falsely portraying the activity as risk-free and legitimate.

Promotional materials were disseminated in multiple languages. Many advertisements employed highly deceptive creative tactics, including AI-generated deepfakes impersonating public figures such as Cristiano Ronaldo, fabricated news segments styled after reputable broadcasters, fake transaction notifications, and manipulated app interfaces. The ads frequently mimicked trusted brands such as PayPal and Revolut.

These materials were identified as manipulated or synthetic based on a combination of visual artefacts, inconsistencies with verified footage of the public figures concerned, and pattern-based classification by Debunk.org’s analytical tools.

Despite being presented by InOutGames as a business-to-business gaming product, Chicken Road operated without valid authorisation in the European Union. The company held only a Curaçao licence and lacked any recognised EU gambling licences or Meta platform approvals, according to publicly available data.

Debunk.org submitted **381 reports to Meta**, which led to the removal of **6,312 pages and 1,466 million advertisements**. The ads reached **over 1,169 billion users across Europe**. This figure is derived from Meta Ads Library estimates for the ads removed in connection with this scheme and should be interpreted as an indicative estimate rather than an exact count of unique individuals.



Individual reports submitted

381

**Submitted by Debunk to Meta*

Facebook pages removed

6,312

**As a result of the reports*

Total reached

1.169 B

**Across Europe*



Advertisements removed

1.466 M

**As a result of the reports*

Inactive
Library ID: 743607981618005
Aug 2, 2025 - Aug 8, 2025
Platforms

This ad has multiple versions

ChickenRoad LT
Sponsored
Library ID: 743607981618005

Kada paskutinį kartą gavai
NEMOKAMAI?
BE RIZIKOS BE ĮMOKOS

+5€

CHICKEN ROAD
ISBANDYK DABAR

GAMESWORLD.ONLINE
★★★★★ (4.9/5) [Learn More](#)

This ad has multiple versions 1 of 3

STIRILE PRO-TV O MAMĂ CARE TOCMAI A NĂSCUT A CĂȘTIGAT 17 MILIOANE DE EURO ÎN NOUA APLICAȚIE CREATĂ DE MARK ZUCKERBERG

STIRILE PRO-TV O MAMĂ CARE TOCMAI A NĂSCUT A CĂȘTIGAT 17 MILIOANE DE EURO ÎN NOUA APLICAȚIE CREATĂ DE MARK ZUCKERBERG



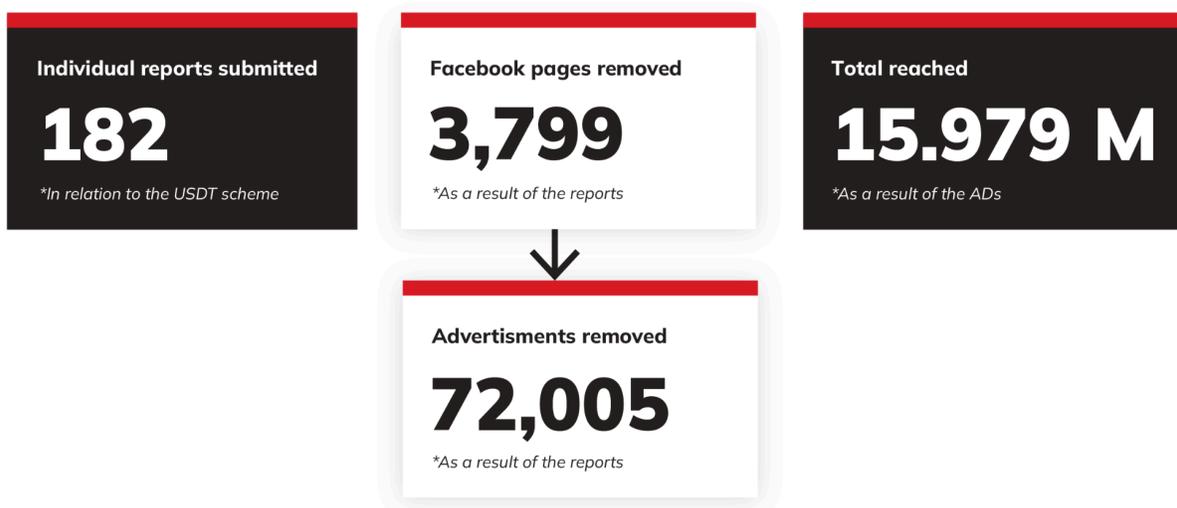
- **The persistent investment fraud scams and the case of USDT**

Some of the most persistent scam schemes analysed by Debunk.org are **investment fraud operations**, in which advertisements systematically promise rapid, effortless, and risk-free financial gains. These campaigns often claim that profits can be generated automatically, with no active involvement from users, frequently invoking the use of artificial intelligence, automated trading systems, and algorithmic “bots” to create a false impression of technological sophistication and reliability. The advertising language is deliberately deceptive, emphasising exceptionally high, immediate, and guaranteed daily returns, and framing the schemes as exclusive opportunities for fast wealth accumulation.

One of the notable schemes is the USDT, a scheme promoting fraudulent cryptocurrency-based investment platforms. The ads found by Debunk promised highly unrealistic financial returns, including claims of earning hundreds of USDT per day from small deposits. The schemes combined exaggerated profit projections, recruitment-based earnings structures, misuse of well-known brand names, and psychologically manipulative marketing techniques.

In addition, the schemes systematically exploited narratives of passive income and financial automation, frequently advertising the use of supposed “AI trading bots” or “advanced staking plans” that allegedly generate profits without risk or effort. Users are introduced to tiered “VIP levels”, loyalty bonuses, and multi-level affiliate or team-based commission structures, which reward recruitment and further financial contributions.

Debunk.org submitted **182 individual reports** on the USDT scheme, which resulted in the removal of **3,799 pages** and **72,005 advertisements**, reaching an **estimated audience of 15,979 million**.



8. Conclusion

This report demonstrates both the potential and the limits of a Trusted Flagger organisation under real-world reporting volumes. Debunk.org was able to identify large-scale scam and fraud networks, highlighting the essential role such tools play in effective enforcement. Despite a high level of takedowns, Meta’s response times remain inconsistent week to week, raising questions under the DSA about what constitutes a “reasonable” response timeframe.

9. Demonstration of Independence pursuant to Article 22(2)(b) DSA

To ensure compliance with the independence requirement set out in Article 22(2)(b) DSA and guarantee that its trusted flagger activities are carried out impartially and autonomously, trusted flagger functions are organised as a distinct and clearly delineated project stream within Debunk.org. All analytical work is conducted in teams and subject to internal peer review by analysts, and if necessary, reviewed by external lawyers.

All notices concerning potentially illegal content are assessed independently, in accordance with established internal procedures and documented methodological standards. Individuals performing trusted flagger functions are not employees of any online platform (VLOP), nor of



consultancy firms working for such platforms, and have no affiliation with platform personnel.