

META ADS : A PILL HARD TO SWALLOW

Meta is still failing to moderate its advertisement ecosystem.

Credits:

Research & Report: Dr. Paul Bouchaud - Al Forensics

Report: Salvatore Romano, Raziye Buse Çetin, Marc Faddoul - Al Forensics

Report: Karla Pajares Sangay, Jinyu Liu - Participants in the DSA Ad Repository Data

Sprint

Graphic & Brand Design: Denis Constant / Ittai Studio https://ittai.co/

All other content (c) Al Forensics 2024

Email: info@aiforensics.org
Social Media: Linkedin | Bluesky

This Al Forensics research was funded by core grants from <u>Open Society</u> <u>Foundations</u>, <u>Luminate</u>, and the <u>Limelight Foundation</u>.

Context:

This report has been authored during the <u>DSA Ad Repository Data Sprint</u> organized by Sciences Po and the Open Institute for Digital Transformations having taken place between February 24th and February 27th, 2025.

Disclaimer:

This report presents analyses and conclusions based on publicly available data as of February 2025. While we have made every effort to ensure accuracy through rigorous documented methodology, all findings represent best estimates with inherent uncertainties.

Each analysis section details its specific methodology and limitations. Sporadic mistakes may have arisen from manual annotation. This assessment reflects current information and may be updated as new data emerges.

When referring to ads as "scams" or "fraudulent" ads, we do so by leveraging documented patterns of misleading practices, yet this analysis <u>does not</u> constitute a legal determination, as judicial proceedings remain ongoing and such assessments lie outside both our expertise and the scope of this report.

Table of Contents

Table of Contents	3
Executive Summary	3
Introduction	
Data Collection	6
Near-Duplicate Detection	6
Exploratory Analysis	7
Results	9
I - Weight loss	9
- "Weight loss industry"	9
"Not just a question of aesthetics"	
Recent Activity	
II - Join Inflammation	
III - Diabetes	16
IV- Prostate	17
Prostatitis	17
Tomato	18
ProsaCare	18
V- Parasite	19
VI- Erectile Dysfunction	21
Discussion	23
Advertising Standards	23
Platform Features Review	25
Beneficiary & Payer	25
Display Link	26
Page Name Impersonation	
A final example	
Potential DSA Infringements	
Conclusion	30

Executive Summary

- Since the enactment of the Digital Services Act (DSA) in August 2023, Meta
 has approved over 46k advertisements containing unapproved drugs and
 deceptive health claims. Together, they appeared on the screens of European
 users over 292 million times.
- Those ads go against at least 15 of Meta's Advertising and Community Standards provisions, such as misleading health claims, celebrity deepfakes, and impersonation of news outlets.
- Meta's failure to curb these fraudulent ads appear to contravene to the mitigation of systemic risks mandated by the DSA Article 34, from the facilitation of distribution of illegal content to foreseeable negative impacts on public health.
- The promotion of unapproved medical products, impersonation of public figures, and widespread use of deceptive landing pages further suggest breaches of additional European and national laws on consumer protection, unfair commercial practices, and pharmaceutical advertising.
- Meta's self-disclosure mechanisms for listing the "beneficiary" and "payer" and the option to display a deceptive URL were commonly abused. These features enabled large-scale scams and phishing campaigns, underscoring the inadequacy of current transparency safeguards.

Introduction

Building on previous investigations conducted at AI Forensics on the Meta Ad Library, this report focuses on health-related scam advertisements. We identify over 46 thousands advertisements, approved and delivered by Meta to european users over 292 millions times, such products typically promoting dubious drugs or "cures" not approved by national or European health organizations.

<u>Our previous work</u> characterized Meta's poor moderation of undeclared political advertisements, revealing an almost complete reliance on advertisers' self-declarations. This systemic shortcoming was exploited by threat actors to disseminate thousands of pro-Russian propaganda ads, all approved by Meta and delivered over a hundred million times to European users of Meta's products. Similarly, we identified massive scam networks attempting to lure users into fraudulent investment platforms, operating across multiple countries since at least the enactment of the Digital Services Act (DSA), and remaining active as of this report's completion.

The severity of these findings motivated the European Commission to open, in April 2024, <u>a formal proceeding</u> against Meta for suspected violations of the DSA.

Following the publication of <u>Meta's Risk Assessment under the DSA</u>, we documented misleading statements in which Meta claimed to automatically review ads against their Advertising Standards before launching on Facebook or Instagram. However, <u>we identified thousands of advertisements featuring explicit pornographic content.</u>
When we uploaded these identical images on Instagram and Facebook as regular posts, they were automatically removed for violating the platform's nudity policy. This simple experiment demonstrates that Meta *does* possess the technical capability to detect such content but has *chosen* not to apply these safeguards to advertisements, their primary revenue source.

We wish to acknowledge that these investigations, including the present study, were made possible through Meta's implementation of a public advertisement repository as mandated by Article 39 of the DSA. Meta's provision of a functional and relatively comprehensive advertisement library and API demonstrates a higher transparency standard compared to industry peers, as documented by the Mozilla Foundation.

This report illustrates how researcher data access functions as an effective mechanism for platform accountability through public scrutiny. We urge all platforms to fulfill their data provision requirements, with the advertisement repository representing an enforcement priority.

Methods

Drawing extensively from the Meta Ad Library, this report assesses the prevalence of alleged fraudulent materials distributed through advertisements on Facebook and Instagram, along with Meta's moderation practices.

Data Collection

Following the Digital Services Act implementation, Meta expanded in August 2023, their ad repository to include <u>all advertisements targeting EU audiences</u>. The Meta Ad Library is accessible through both a web interface and API, enabling programmatic analysis of the repository.

Meta's implementation of full-text search capabilities in their ad library represents a significant transparency measure, distinguishing it from other major platforms where such functionality is <u>often unavailable</u>.

One limitation is that only advertisements having concluded their distribution within the last year are available, in literal compliance with DSA Article 39.1. Our analysis will leverage a dataset of over 450 million advertisements, continuously collected through the dedicated API since the enactment of the DSA in August 2023.

Near-Duplicate Detection

To navigate this gigantic dataset, we focused on instances where identical or nearly identical text¹ was published by multiple advertiser pages. The rationale behind this heuristic, <u>employed in our previous studies as well</u>, is that while advertisers may have legitimate reasons to republish identical advertisements (such as varying targeting parameters), this pattern becomes highly suspicious when tens, hundreds, or thousands of distinct pages publish identical advertising content.

This approach effectively surfaces potential coordinated advertising campaigns that warrant further scrutiny. To ease manual annotation, we solely consider texts published by *at least* five different pages and we translate them into English through Facebook's <u>M2M100</u> translation model, chosen for its open-source weights and state-of-the-art performance.

_

¹ When we refer to ads "text" throughout this report, we specifically mean the "ad_creative_bodies" field returned by the Meta Ad Library API.

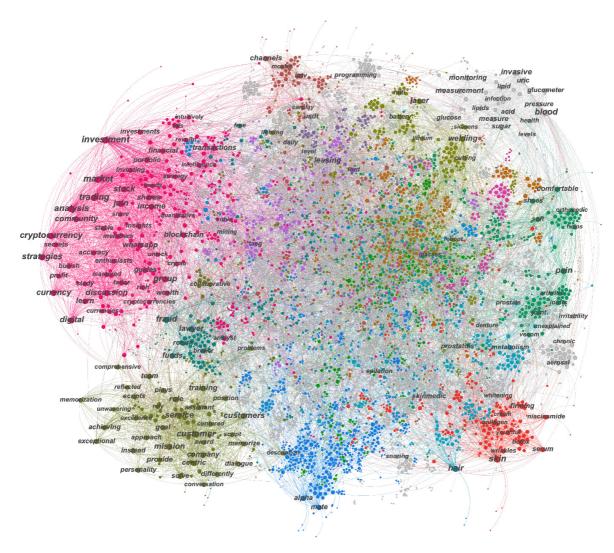
Exploratory Analysis

We identified 226k unique texts published by at least 5 different pages, totaling 7.0 million advertisements with 139.9 billion impressions. While some instances represent legitimate advertising practices—such as Temu's campaign of hundreds of thousands of identical-text advertisements distributed across thousands of different pages, reaching billions of impressions—many others warrant investigation.

To systematically explore this potentially fraudulent content, we first conducted an exploratory semantic analysis.

Specifically, we extract words (minimum 4 letters) appearing in at least 25 different unique texts and compute pairwise chi-square statistics. For two words, the chi-square statistic measures the difference between the observed co-occurrence frequency and the expected frequency under independence. This metric, also used in the <u>Cortext</u> platform, assesses word similarity.

We represent similarity as a network where nodes represent words and edges connect words that co-occur more frequently than expected (defined by a chi-square statistic threshold). After spatialization through ForceAtlas2, words with similar usage patterns cluster together, revealing topical structures within duplicated advertisements.



Communities of words frequently used together are color-coded, with node size corresponding to the reach of advertisements containing these terms.

We observe diverse topics within our pool of suspicious duplicate advertisements. A large cluster promotes dubious cryptocurrency trading and investment strategies.

Notably, the right side of the network reveals substantial health-related subtopics, ranging from skincare products and serums to treatments for arthritis, prostate issues, weight loss, and blood pressure regulation.

In the subsequent analysis, we select a handful of keywords associated with various medical conditions and report the most duplicated and widely distributed advertisements.

Importantly, we neither seek nor claim to be exhaustive; given the enormous volume of fraudulent content approved by Meta, we aim solely to demonstrate the pervasiveness of this issue through various examples.

Results

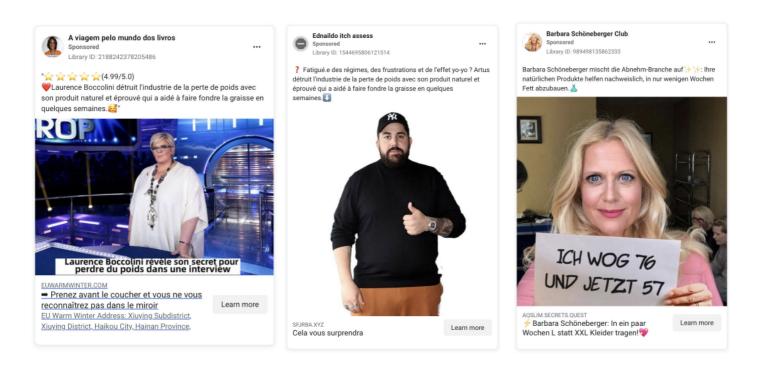
I - Weight loss

"Weight loss industry"

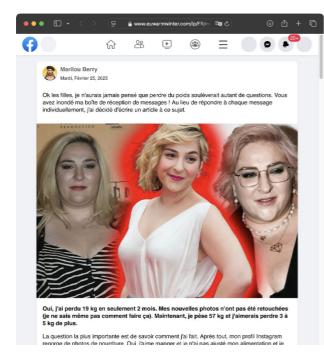
By searching for advertisements claiming to go against the established "weight loss industry with" their product, we find **4,013 advertisements** matching the pattern, with 17 unique text variations, delivered at least **48.8 million times** by Meta to users in France, Germany, and Spain.

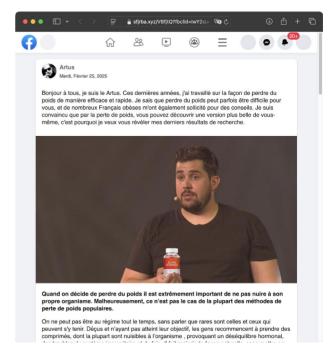
The advertisements typically follow the pattern: "[Celebrity name] destroys the weight loss industry with its natural product and proves that it helps to melt fat in a few weeks" and other analogous variations.

Overall, they target Marilou Berry, Artus, Laurence Boccolini, and Dr. Michel Cymes in France, Barbara Schöneberger in Germany, and Paula Echevarría in Spain.



The advertisements refer users toward phishing pages taking the appearance of a Facebook post promoting the weight loss product.





Snowballing and searching for the name of "Marilou Berry", a french actress, in duplicated advertisements (excluding the previously identified ones), we find **973** additional ads discussing weight loss, for a total reach of **15.7 million**. We observe additional misleading practices such as redirecting users toward fake spoofed news outlets, in this case "Le Figaro" in French.





For "Artus", a French comedian, we find **403 new ads** for a total reach of **4.7 million**, containing a handful of variations of "Artus reveals the secret of weight loss."





For "Laurence Boccolini", a French TV host, we found **312 new ads** for a total reach of **4.1 million**, related to recipes for weight loss.





"Not just a question of aesthetics"

We find an additional pattern, translated in French, Spanish, German and Italian, used in over **1897** advertisements, delivered by Meta over **23.2** millions times:

"Weight loss is not just a matter of aesthetic or appearance. It is an important step towards the improvement of the holy and quality of life. When we take care of our weight we reduce the risk of many diseases such as diabetes, cardiovascular diseases and hypertension. Maintaining a healthy weight helps equally to strengthen the immune system to improve sleep and general wellbeing ..."







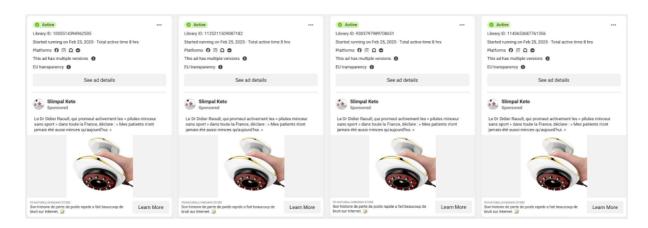




Recent Activity

Far from being shortcomings of the past, while we were performing the present analysis, numerous fraudulent advertisements were actively distributed by Meta to European users.

For instance, advertisements were approved and distributed on February 25th, 2025 for "weight loss pills" redirecting to a fake spoofed news outlet with a fake interview of decried French microbiologist Didier Raoult, with fake certificates of "conformity" supposedly issued by the World Health Organization.







The drug "Proper Keto" does not appear in the <u>French public drug database</u> as of February 2025.

II - Join Inflammation

Through queries for duplicate advertisements containing "arthritis" and "hernia," we identified **2,637 advertisements** delivered by Meta approximately **43.5 million times**. These advertisements typically presented variations of the text: "Gout, arthritis, arthrosis, osteochondrosis, rheumatism, hernia, sinovite etc. And all these problems are cured in a very simple way a cleaning of the joints..."

This content was translated, with minor variations, into Spanish, Italian, Portuguese, Croatian, and Bulgarian. These advertisements predominantly targeted elderly users (50-65+), with numerous instances utilizing fabricated television segments to promote unapproved pharmaceutical products.







III - Diabetes

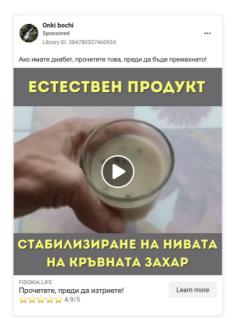
We find **2,480 advertisements**, disseminated by Meta across the EU over **17.7 million times**, all following the pattern "If you have diabetes, read it before it is removed" translated into national languages with slight variations.

Interestingly, many of the pages running such ads follow the naming convention observed and used in the pro-Russian propaganda Doppelganger operation, reported by <u>Reset Tech</u> in October 2023 and in <u>our previous report</u> in January 2024. Such pages are still active as of February 2025.













IV- Prostate

Prostatitis

Looking for "prostatitis" we find 9,656 advertisements, for a total reach of 42 millions, run in Romania, Spain, Bulgaria, Poland and Italy, containing 15 unique variations of the following text:

"How I helped more than a million [Nationality e.g. Spanish] men get rid of prostatitis in years with this method of [a pretended cure e.g. roses, bicarbonate]. This indicates the true cause of prostatitis that has nothing to do with bacteria and with age eating junk food or lack of movement but is actually linked to a lack of individual cells that inhibit the prostate..."







Tomato

We find **15,945** advertisements with 4 unique texts running in English, French, Spain, for a total of **1.9** millions reach, with slight variation of the following:

"I could no longer bear taking more medication and visiting the doctor for my prostate That is when I ended up watching a video on the internet that talked about a trick with tomatoes that helped keep the prostate healthy in just a few days I saw many friends talking about this so I decided to watch the video and then followed all the steps with the tomato to keep my prostate healthy It was amazing We really spend money on medications full of side effects"





ProsaCare

We find **202** ads promoting a pretended drug "ProsaCare" delivered by Meta **5.6 million times**, with 3 unique text variations in German and Dutch. The leading webpage is disabled as of February 2025 and signaled as a probable scam by users on Trustpilot.

V- Parasite

We find **3,164 advertisements**, disseminated by Meta across the EU over **18.6 million times**, discussing parasitic infections, enumerating symptoms, and ultimately promoting a pretended cure.

For instance: "Warts and papillomas are caused by parasites in the body. There is a new way to get rid of parasites at home."













Among the advertisement materials, we can mention a fake TV segment from France24 spoofed with deepfakes of journalist and contested French microbiologist Didier Raoult. The fake TV segment claims unsanitary conditions lead to parasite infections and redirects users toward a marketplace to buy pretended "certified and proven" drugs. [see advertisement].







VI- Erectile Dysfunction

We find **1,925** advertisements, shown by Meta to EU citizens over **16.4** million times, following the pattern "How to get rid of erectile dysfunction once and for all using cheap and natural remedies" translated into national languages with slight variations. When explicit targeting options are set, they typically tend to focus on older men (40-65+).

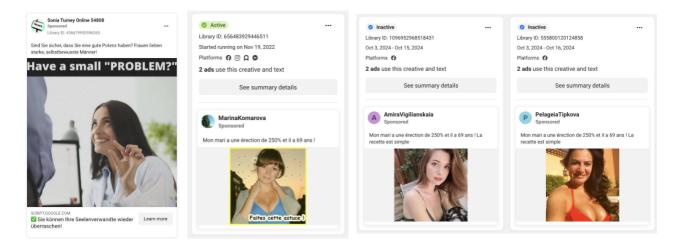


We found **564** ads in Italian, approved by Meta and delivered over **19.2** million times, following the two patterns: "What is the fastest way to cure erectile dysfunction in two weeks? The 7 most effective drugs for erectile dysfunction in 2024!" and "Are you looking for a safe and effective way to improve your erectile dysfunction?"





We find **469** ads in German delivered by Meta **4.2** million times, following the pattern "Are you sure you have good potency? Women love strong, confident men!".



Additionally, we find **99 ads** in French shown by Meta at least **2.2 million times**, following the pattern "*My husband has a 250% erection and he is 69 years old! The recipe is simple*" with a photo of an appearing woman. The activity of such ads dates back to at least 2022, with an advertisement appearing as still active on the Meta Ad Library in February 2025.

We found **808** ads in German shown by Meta **10.6** million times, for a product "approved by doctors" that supposedly "i) get a larger penis ii) Last more than 2 hours iii) Increase your masculinity."

Similarly, we found **715 ads** in French shown by Meta at least **9.3 million times** to "Increase the size and volume of your penis at least 40% compared to its previous size." [we blurred sexually explicit material in the image below.]

As a last example, we found **357 ads** in Romanian shown by Meta at least **4.3 million times** with the same text: "From the first use, you will feel a surge of strength in your penis. And you can have sex for 2 hours without stopping"







Discussion

Advertising Standards



After examining Meta's policies, we identified multiple violations of their advertising guidelines and community standards against which advertisements are purportedly reviewed prior to distribution by Meta. From the previously documented advertisement, we can identify potential violations of the following 15 policies:

According to Meta's policy on Fraud, Scams and Deceptive Practices, ads can not:

- Employ click-bait tactics in a health context, such as the use of sensational language that makes exaggerated or extreme health claims.
- Promise specific health-related results in specific time with no qualifying or disclaimer language.
- Promote claims to cure, heal, or eliminate [..] diabetes [among other illnesses].

According to Meta's policy on Unacceptable Business Practices, ads can not:

- Use the image of a famous person and misleading tactics in order to bait people into engaging with an ad.
- Use deceptive or exaggerated claims about health-related benefits of a product or service to mislead people into purchasing or sharing sensitive information.
- Use deceptive or exaggerated claims about the success of a product or service to mislead people into purchasing or sharing sensitive information.

According to Meta's Health and Wellness policies, ads cannot:

- Promote weight loss products and services that contain distasteful messaging that could make people feel negatively about the way they look, exploit insecurities to conform to certain beauty standards, reinforce negative or unhealthy body images, or feature body-shaming
- Promote weight loss products or services depicting side-by-side comparison after the use of a product
- Promote sexual arousal products that focus on sexual pleasure or enhancement, such as [...] non-surgical genital enhancement products
- Promote genital procedures or surgeries focused on sexual pleasure, such as G-spot augmentation, male genitalia enlargement

According to <u>Meta's policy on spam</u>, one cannot share:

- Deceptive URLs mimicking the features or functionality of their services.
- Landing page impersonation of well-known websites, domains or brands.

According to Meta's policy on Inauthentic Behavior, one cannot:

 Misrepresent themselves on Meta's services, use fake accounts or engage in behaviors designed to enable other violations under their Community Standards

According to Meta's policy on Evading Enforcement practices, advertisers cannot:

 Create and run the same or similar policy-violating ads across multiple business assets to evade their review processes

According to <u>Meta's Vice President, Global Policy Management statement</u>, content will be removed:

"If It has been edited or synthesized [..] in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say. And is the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic."

Our investigation shows that Meta fails to enforce their established advertisement and community standards. The extent of these enforcement failures raises significant concerns regarding the efficacy of Meta's risk mitigation protocols.

Platform Features Review

While we acknowledge that online scam may be a highly adversarial space with some malicious groups evolving their tactics to evade detection, we emphasize how plain, un-obscured, policy-violating content has been approved, at scale, by Meta.

Moreover, features offered by the platform to its customers (i.e., advertisers) facilitate such fraudulent activity, raising compliance concerns.

Beneficiary & Payer

As to comply with DSA article 39.2.b, <u>Meta implemented an option</u> for advertisers targeting the EU to declare the person or organization who is benefiting (beneficiary) from the ad and, if different, paying for it (payer).

The mechanism put in place by Meta relies on advertisers' self-disclosures —as acknowledged in <u>Meta's DSA Risk Assessment</u>, and emphasizing that advertisers "are responsible for ensuring that this information is complete, accurate and up-to-date for each ad that [they] submit to Meta"



Screenshots of the Meta Ads Manager, self-disclosure of beneficiary and payer

Within the previously identified fraudulent advertisements, beneficiary and payer information tended to be either automatically generated or nonsensical gibberish (e.g., "vdfvbdf", "twovfdqu", "mm13").

While Meta does have access to payer names for credit card transactions and has <u>established processes to verify advertiser IDs</u> for self-declared political advertisements, relying on apparently unchecked self-disclosure appears to conflict with DSA Article 39's objective to enhance transparency and accountability

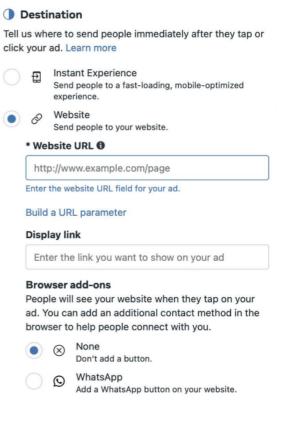
Display Link

Meta offers advertisers the option to display a URL to users that differs from the actual destination where users will be redirected, perfecting the phishing illusion.

Far from a remote risk, we identified fraudulent advertisements displaying links to legitimate news domains such as "Le Monde" in France while redirecting to scam webpages.







Page Name Impersonation

Per Meta's policy, individuals are prohibited from impersonating others to mislead or deceive, and impostor accounts are reportedly disabled immediately upon detection.

However, a simple search of advertiser pages using names of health-related public figures (e.g., Didier Raoult or Michel Cymes in France) reveals numerous impostor accounts that have published advertisements. These advertisements exhibit fraudulent characteristics, employing audio deepfakes of the impersonated celebrities within simulated news TV segments to promote pharmaceutical products.

It appears that Meta's systems fail to detect such fraudulent activity, despite <u>public</u> <u>commitments</u> to do so.

Advertisers



Pr.Didier Raoult

3 545 follow this · Medical & Health



Dr. Didier Raoult

38 follow this · Medical & Health



PROF. Didier RAOU

3 follow this · Advertising Agency



PROF. Didier RAOU

2 follow this · Advertising Agency



Didier Raoult

31 follow this · Women's clothing store



Dr Didier Raoult

148 follow this · Farm







Michel Cymes

3 follow this · Producer



Michel Cymes

5 follow this · Public Relations Agency



Michel Cymes

2 follow this · Public Relations Agency



Michel Cymes

3 follow this · Website



Dr Michel Cymes

11 follow this · Medical & Health



Dr. Michel Cymes

6 5 follow this · Health/beauty



Dr. Michel Cymes

225 follow this · Business Service



Dr. Michel Cymes





A final example

Our previous observations can be crystallized through the following advertisement:

- Run by an impersonator page titled "Didier Raoult" using his image as profile picture
- Promoting a pharmaceutical product against arthrosis targeting the elderly (55-65)
- Use a spoofed media outlet with audio deepfake
- Redirecting users to a counterfeit news site mimicking French newspaper "Le Monde" to promote, via fabricated interview with Didier Raoult, said pharmaceutical product
- Presenting a lottery offering discounts on the pharmaceutical product, a practice explicitly prohibited under the European Directive 2001/83/EC on the <u>Community code relating to</u> <u>medicinal products for human use</u>
- Soliciting users' personal data (name and phone number)
- Promoting a drug absent from the French public database of authorized pharmaceuticals



Se connecter S'abonner dès 5,49 €/mois



"SI L'ON VOUS DIT
QU'APRÈS 50 ANS VOUS NE
POUVEZ PLUS SOIGNER
VOTRE DOS ET VOS
ARTICULATIONS, SACHEZ
QUE C'EST UN MENSONGE !"

CULTURE V LE GOÛT DU MONDE V SERVICES V Q

@ ¿

Potential DSA Infringements

Through the Digital Services Act (DSA), the European regulator acknowledged that, on account of their scale, very large online platforms' advertising systems can pose potential risks and as such require public and regulatory supervision.

Specifically, <u>recital 95</u> of the DSA mentions the "emerging risks brought by the distribution of advertising online, for example in relation to illegal advertisements or manipulative techniques and disinformation with a real and foreseeable negative impact on public health".

Our results revealed systemic shortcomings in Meta's advertisement moderation. While the actors behind the documented fraudulent activity are using adversarial methods to scale their operations, these methods remain rudimentary and fall well within what one could expect a platform such as Meta to detect and act upon.

Going further, <u>recital 79</u> of the DSA specifies that very large online platforms such as Meta should assess and mitigate the systemic risks stemming from the use and misuse of their services.

The systemic risks outlined in the regulation article 34.1.d explicitly mention the dissemination of illegal content through platform services, as well as actual or foreseeable negative effects in relation to the protection of public health and serious negative consequences to individuals' physical and mental well-being.

The advertisements we identified have all been reviewed and explicitly approved by Meta's moderation systems. Their fraudulent nature ranges from selling drugs not approved by targeted market authorities —an illegal practice under French law for instance—, deceiving the elderly with promises of miracle cures for their illnesses —condemned by the European Regulation on Unfair Commercial Practices, to misleading practices through the use of celebrity deepfakes and news outlet impersonation; a practice that violates the Code of Practice on Disinformation, to which Meta is a signatory.

Far from merely resulting from the actions of bad actors, we identified platform features offered by Meta to its customers (i.e., the advertisers) whose misuse appears evident and expected. We document their use in fraudulent practices.

Among these, the disclosure of the natural or legal person on whose behalf an advertisement is presented—a transparency feature <u>required per the DSA</u>—was implemented by Meta as mere self-disclosure without apparent verification, which

led to false declarations by advertisers involved in fraudulent activity. This practice was arguably predictable and could have been mitigated by the platform.

Additionally, Meta offers advertisers the possibility to choose a link appearing on ads that differs from the actual destination where users will be redirected. This feature is exploited by fraudulent actors to spoof news outlet websites, another predictable practice that could have been mitigated by the platform.

Finally, we found numerous advertiser pages impersonating celebrity doctors in health-related advertisements to lure users into buying products. Far from employing elaborate obfuscation techniques, these pages plainly use the name and image of the impersonated figure. Verification mechanisms for advertiser page names could reasonably be expected.

Conclusion

This report highlights that Meta's Ad repository serves as a robust transparency tool, enabling third-party scrutiny. Following our previous report, this analysis reveals Meta's continuous failure to effectively moderate its advertising ecosystem.

Though not comprehensive, by detecting highly duplicated content containing a handful of keywords of interest, we identified at least 46 thousand health-related scam advertisements, approved and delivered by Meta, over 292 million times to its European users.

The scale of the fraudulent activity, combined with platform features that facilitated scammer operations, raises concern with respect to Meta's responsibility in such deceptive practices.