ADVERTISING AND THE ONLINE SAFETY ACT 2023

Prof Lorna Woods OBE and Dr Alexandros Antoniou

Summary of key points

- Online advertising remains fraught with risks (from scam content to ad misplacement) and current regulation struggles to address increasingly opaque and automated delivery systems susceptible to misuse.
- The Online Safety Act 2023 (OSA) marks a partial shift by extending duties to user-to-user (U2U) and search services for policing fraudulent ads, but the scope of these responsibilities remains unclear.
- Ambiguities in the OSA's definitions, particularly around 'user-generated content', 'users' and 'services', create uncertainty that weakens the regime's enforcement potential in this context.
- Whether advertisers are considered 'users' affects whether general safety duties apply to ads, and this lack of clarity impacts the reach of illegal content provisions across different types of services.
- Search services remain largely exempt from the OSA's broader duties on illegal content in paid-for ads, creating enforcement asymmetries between regulated services despite equivalent exposure to harm.
- The Online Advertising Programme (OAP), originally intended to fill regulatory gaps left by the OSA, has thus far delivered minimal progress, focusing narrowly on content rather than systemic ad delivery flaws.
- The OSA introduces vital duties but falls short of delivering a comprehensive regulatory framework, leaving some gaps in coverage, enforcement, and oversight of third-party ad delivery systems.

Introduction

Advertising is a significant element of the online environment, yet concerns persist regarding its content, particularly around scams and fraud, and the blurred line between organic content on the one hand and commercial/ sponsored material and the role of influencers on the other. Additionally, there are growing apprehensions that platforms hosting and delivering these ads do not respond to complaints about them. Further issues arise from ad placement, including instances where ads appear alongside terrorist content, child abuse material or misinformation and made-for-advertising (MFA) websites (including e.g., obituary scams). Recent research underscores these risks, alongside questions about whether ads even reach their intended audience, given techniques like bot traffic, pixel stuffing and ad stacking. The largely automated and opaque ad delivery system is susceptible to exploitation by bad actors, including fraudsters, unscrupulous publishers, or even opportunistic intermediaries, seeking to game the system's vulnerabilities.

In this piece, these placement and visibility issues, often involving manipulation of ad placement systems and improper ad delivery, can be referred to as 'ad fraud'. We do not include in this category concerns related to the targeting itself (e.g., bias in the algorithm and data protection issues). By contrast, we term any advertisements containing scams and misleading content as 'fraudulent ads' - a term that is a broader than the Online Safety Act (OSA) definition of fraudulent ads. Fraudulent ads form a sub-set of adverts that contravene the criminal law.

In the light of these issues, we review the existing regime. This piece first outlines the existing advertising framework, which focuses principally on ad content. It then considers the extent to which the OSA, which addresses social media and search content delivery systems, tackles both ad fraud and fraudulent ads, as well as considering briefly the approach to criminal ads more broadly. We highlight drafting ambiguities in the OSA that cast doubt on whether the policy objective of excluding both ad delivery and ad content from the regime (with the exception of fraudulent ads) is fully achieved. This interpretation opens the possibility of the application of a wider range of priority content rules to ads on social media, while preserving protections against fraudulent ads. The position regarding delivery systems and ad fraud is less clear. Finally, this piece questions the role of the Online Advertising Programme (OAP) - launched by the Johnson administration to review the whole paid-for online advertising ecosystem, and prolonged under Labour - and the extent to which it will fill any gaps left by OSA, specifically around ad fraud.

1. Regulation of content: the existing regime

Advertising content is regulated by the Advertising Standards Authority (ASA), with non-broadcast adverts falling under the CAP code. This includes the content of adverts provided through social media or other online ad delivery services, and marketing communications in user posts by celebrities and influencers. The CAP Code covers a range of issues – including harm and offence, adverts directed at children, and misleading ads. It also sets out specific rules about certain products or services (e.g., medicines, foods/ food supplements and gambling as well as financial products). A central principle is that marketing communications should be 'obviously identifiable as such'. The ASA describes its aims as ensuring the advertising is 'legal, decent, honest and truthful'.

The system is principally self-regulatory (although some rules reflect legal obligations and may also fall within the remit of Trading Standards or the Competition and Markets Authority (CMA) as a backstop). The ASA does not have any powers to fine non-compliant advertisers, nor to compel their compliance with their rules, relying instead heavily on industry buy-in and a 'name and shame' approach. For certain consumer protection rules, the ASA can make referrals to the Trading Standards authorities.

One persistent issue pertains to influencers failing to clearly disclose when their content is advertising (i.e., when it has been provided in exchange for payment, gifts, or other forms of value). Despite the ASA introducing guidance in 2018 and updating it in 2020, a 2021 monitoring report showed 'unacceptable levels' of non-compliance. Measures to encourage compliance for traditional formats seemed not to affect influencers. In response, the ASA introduced in June 2021 a dedicated webpage to 'name and shame' influencers 'routinely failing' to comply with advertising rules regarding disclosing marketing intent. It started taking out social media ads to alert users to specific non-compliant influencers. Following an investigation by the CMA in 2018, which also demonstrated the problems in this area, the CMA, ASA and Ofcom jointly published three sets of guidance in 2022 for social media companies, brands and influencers respectively to tackle 'hidden ads'. The ASA followed up with a third edition of its influencer guidance in 2023. Despite these efforts, effective enforcement or encouraging compliance in this context has proved difficult, particularly as the rapid growth of short-form audio-visual formats has made it even easier for promotional content to blend seamlessly and appear increasingly organic.

Scam adverts have been a persistent concern for a while and were specifically identified by the 2017 Green Paper, and recent research shows that under-18s are exposed to them. Some online paid-for adverts link through fraudulent content often featuring fake celebrity endorsements or fabricated stories (see e.g., Martin Lewis/ Money Saving Expert). In one instance, paid-for ads

featured fake nudes of BBC presenter Naga Munchetti. *The Guardian* also reported on scammers operating at scale using fake celebrity ads, with illicit gains totalling £27 million. While victims were targeted globally, about 45% were UK-based. Which? suggests that four common types of scam ads are: fake celebrity endorsements; AI-generated deepfakes; travel scams impersonating legitimate companies with unrealistic low prices; and fraudulent brand sales exploiting established reputations. In its OAP consultation, DCMS noted that many larger platforms offer 'self-service' advertising buying services, where there is little vetting for advertisers:

> *As a result, bad actors often operate with relative impunity, using online advertising as a means to perpetrate fraud or advertise other illegal or legal but harmful products and services, with limited oversight.*

The ASA (working with the Internet Advertising Bureau) launched in June 2020 a Scam Ad Alert service, allowing users to report fraudulent ads to the ASA. Once submitted, these reports are shared with organisations participating in this scheme, including Google; Meta (Facebook/ Instagram); Taboola; Outbrain; Microsoft; TikTok; Yahoo; Snap; Twitter; Amazon Ads; Sizmek Ad Suite; RevContent; Index Exchange; Clean.io; Reach; and the Media Trust. Ads and accounts may be taken down and added to a blocklist by participating services. Platforms' response rate to notifications from the ASA are far from swift: the advertising authority reported (in Feb 2025) that platforms responded to alerts within 48 hours in 71% of cases, while 16% received responses beyond this time-frame, and 13% went unanswered. Moreover, there seems to be no mechanism for ads reported directly to the platform to be fed into this system for action. Reports to the ASA are also shared with the National Cyber Security Centre (NCSC) to be dealt with as part of its work against phishing and scams.

The ASA requires sector-specific regulations to be respected (see e.g., the ASA's work on unregulated investments falling outside of traditional financial regulation). In addition to the ASA, some sector-specific regulators enforce advertising rules in relation to their respective areas of competence – most notably in relation to scams, the FCA. It intervened to ensure the withdrawal or amendment of over 10,000 misleading financial adverts in 2023 and launched targeted actions against so-called 'fin-fluencers' promoting financial services products illegally. The FCA and ASA have a memorandum of understanding outlining their co-ordination and intersecting responsibilities. However, some concerns have emerged (most recently by the Transparency Task Force) about the effectiveness of their regulatory alignment, particularly in cases where firms outside the FCA's perimeter engage in potentially misleading advertising. The scope of the rules here is broader than ads covered by fraud offences.

2. Publication and dissemination of ads: the Online Safety Act 2023

The ASA and the CMA's targeting of social media platforms and other online intermediaries, was far from comprehensive, so the Government recognised the need for a more coordinated approach. Yet, the initial draft version of the Online Safety Bill excluded advertising. Clause 39(2) of the draft Bill expressly excluded 'paid-for advertisements'. This took out of the regime the content of ads delivered by regulated services with a further sub-clause describing paid-for ads as those where the payment goes to the delivery system (clause 39(7)(b)). Arrangements where the payment would go to the content producer would therefore not have been excluded. Thus, influencer marketing content would have been relevant content for the draft Bill. A similar approach could be seen in relation to search services (see clause 134(5) of the draft Bill). Even though influencer content was caught by the draft Bill, fraud offences were not listed as priority offences triggering the regime.

Following a successful campaign, including consumer groups, financial sector organisations and mental health charities, and Parliamentary concerns expressed through pre-legislative scrutiny of the draft Bill, the Government confirmed it would include fraudulent 'paid-for advertisements' in the regime, adding a separate set of duties (Chapter 5) and adjusting some of the main definitions in response. The position regarding influencers remained unchanged. As we demonstrate below, while the policy intent seemed to aim generally to exclude platform- and search-delivered ads from the regime apart from these new specific rules, this series of changes has introduced some ambiguity into the scope of the OSA regime overall.

Our reading of the OSA has identified a number of potential concerns. These are summarised as follows:

- Chapter 5 imposes different obligations in relation to fraudulent ads on U2U services from search services.
- The content to which these Chapter 5 duties about fraudulent ads apply is narrower than the content covered by the duties in the other parts of the Act that apply to U2U and search services. This leaves a potential gap in protections and enforcement relating to advertising.
- Small U2U and search services are not covered by the Chapter 5 duties. This also leads to significant gaps in fraudulent ad protections, including for children.
- Paid-for adverts are included in the U2U (Part 3) provisions but are not included in the search (Part 3) provisions.

- Definitions are unclear in the Act itself: a 'user' (for the purposes of 'user-generated content') could include advertisers uploading content to a service as well as individual users generating their own text or image-based content; the 'service' (for the purposes of sharing the content that is regulated by the Act) could include the advertising delivery service as well as the user interface. Different interpretations of these terms have significant implications for the scope of the duties.
- Under the most expansive interpretation of 'user' and 'service', all advertising linked to Schedules 5-7 comes into scope of the illegal content duties but only for U2U services *as well as* the content listed at section 40 ('fraud etc. offences') coming into the scope of specific fraudulent ad duties for U2U and search services.
- Under the most expansive interpretation, obligations would apply to advertising content on U2U already as the general rules are already in force (or shortly to be in force in respect of content harmful to children).

2a. Chapter 5 rules pertaining to fraudulent ads

The fraudulent ads provisions are found in Chapter 5 of Part 3 of the OSA. The respective duties for U2U and search services differ slightly to account for the distinct nature of each service. In-scope services are required to operate their services using proportionate systems and processes designed to:

- prevent individuals from encountering fraudulent advertising by means of the service (Category 1) or via search results (Category 2a);
- minimise the amount of time that fraudulent advertising is present; and
- swiftly remove fraudulent advertising once they are made aware of it through any means (Category 1) or swiftly ensure fraudulent advertising is no longer available once they are made aware of it through any means (Category 2a).

The rules are triggered only by adverts that relate to certain fraud crimes listed in section 40 OSA. This is a much narrower list of offences than those covered by the general illegal content duties. As with the general illegal content duties, ads that do not satisfy the definitions of these s 40 offences will not trigger Chapter 5 of the regime. This leaves a significant gap in the regime in relation to adverts: for example, adverts for child sexual abuse material would not trigger any obligations under Chapter 5. This makes the question of whether the general duties relating to illegal content apply to the content of adverts more pressing (discussed 2c below).

2b. Application of the Chapter 5 rules

The Chapter 5 rules are imposed only on a subcategory of U2U and search services respectively: Category 1 and Category 2a. However, although relevant regulations to start the categorisation process have recently come into force, we do not know exactly which services will be caught – and so the rules remain unenforceable for now. Generally, it seems that only large services will be caught as Category 1 or Category 2a services - a decision which was controversial given that the Act allowed for services to be included in Category 1 on the basis of risk as well as size. So, scams on small services (even those aimed at children) will not be caught by the regime. Moreover, the Chapter 5 duties are only enforceable from the date Ofcom's Code of Practice (s 41(4)) on implementing these provisions is in force (s 51(7)), which is not expected until 2026. In late March 2025, UK Finance (a trade association for the UK banking and financial services sector) and consumer advocacy group Which? called for swifter action on fraudulent ads in a joint letter to the Government. The timing as yet remains unchanged.

Only those adverts (as defined under ss 38(3) and 39(3) together with s 236 of the Act) are regulated content for the purpose of the Chapter 5 duties:

Cat 1 services	**Cat 2A services**
In relation to Cat 1 services, subsection 38(3) defines a 'fraudulent advertisement' as an advertisement that meets all three of the following conditions:	**In relation to Cat 2A services, subsection 39(3) defines a 'fraudulent advertisement' as an advertisement that meets both of the following conditions:**
***It is a paid-for advertisement:* The content has been placed as part of a paid arrangement, rather than appearing naturally or for free, and the placement of the advert is agreed between the parties to the contract.**	***It is a paid-for advertisement:* The content has been placed or promoted in exchange for payment (i.e., involves a financial transaction to be displayed on the search service).**
***It amounts to an offence specified in s 40:* The advertisement's content is illegal because it falls under the types of criminal fraud offences listed in s 40.**	***It amounts to an offence specified in s 40:* The advertisement's content is illegal because it falls under the types of criminal fraud offences listed in s 40.**
***It is not regulated user-generated content:* The content must be an advertisement in its own right, i.e., created and distributed through paid advertising rather than posted directly by users.**	

In summary, for the purposes of Cat 1 duties, an advert qualifies as 'fraudulent' if it is paid-for and linked to a s 40-listed criminal offence that is not user-generated content under the Act.	**In summary, for the purposes of Cat 2A duties an advert qualifies as 'fraudulent' if it is both paid-for and its content amounts to a s 40-listed offence.**

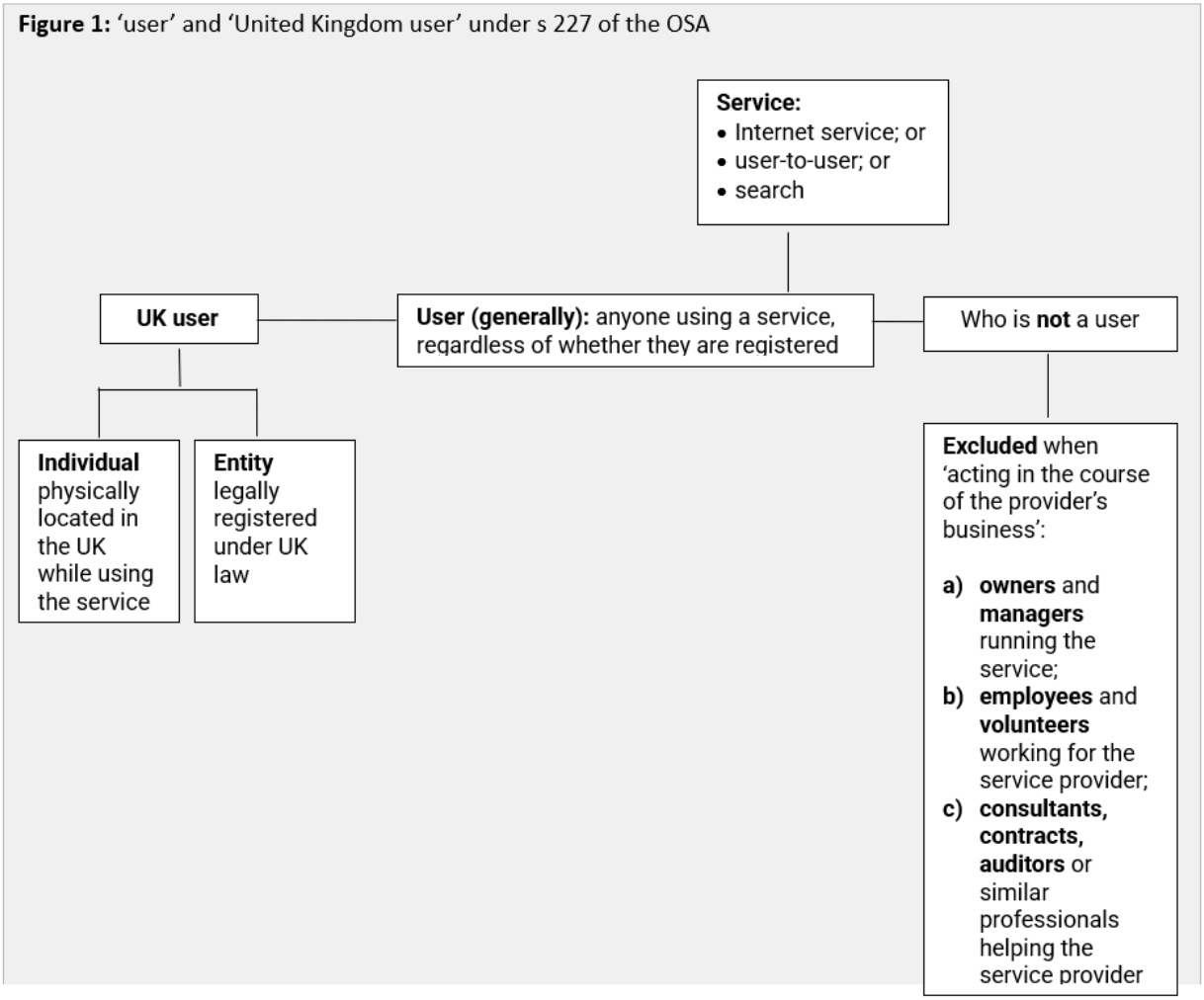
As well as introducing these provisions, the revisions also removed the exclusion of paid-for ads in relation to U2U services but left it intact for search services (see s 57(2)(a) and definition of 'paid-for advertising' in s 236). The implications of this are not clear as we detail below.

2c. Identifying the boundary between the general rules and Chapter 5 rules

The key point of distinction between the general rules and the Chapter 5 rules seems to be the fact that the definition of an advert for the purposes of the specific fraudulent ad duties excludes 'user-generated content' (see the s 38(3) box above). This suggests the general Part 3 duties would apply to what has been termed 'organic content' (which could include some commercial messaging in posts, product placement and the like) produced by 'users' but *not* messages distributed through paid systems fed into the system by advertisers, which would be caught by the Chapter 5 rules instead. This interpretation would match the policy intent noted above (see section 2, p. 4 of this note earlier).

This interpretation depends on the meaning given to 'user-generated content'. The term is defined in s 55(3) and has two core elements: (a) being 'generated directly', 'uploaded to' or 'shared on' the service; and (b) being capable of being encountered by other users. This aligns with the definition of a U2U service in s 3(1) which contains the same two elements. The *second* core element is non-problematic: advertisements are inherently intended for user visibility. The *first* element requires the (direct) supply of the material to the service, which presumably happens when the advertiser places the advert for delivery. So, the question is whether the service referred to is the same one for advertisers and other users. The text does not provide a clear answer, but non-advertiser users encounter organic posts and advertising simultaneously within the same interface – from their perspective there is one service, supplying both forms of content, i.e., organic and advertising content together.

Figure 1: 'user' and 'United Kingdom user' under s 227 of the OSA



One final possibility is that ad delivery clients are *not* considered 'users' of regulated services, and consequently their content is not classified as 'user-generated'. The definition of 'user' (defined in s 227 of the OSA, see Figure 1) - which underpins 'user-generated content' (defined in s 55 of the OSA) - does not seem necessarily to exclude advertisers from being considered users. Specifically, s 227 appears primarily intended to prevent company employees from being deemed users of the service, rather than seeking to distinguish between groups of persons who might use the service in a different way. Likewise, the requirement that a user be in the UK is about the regime's geographical scope, not the nature of the users. Consequently, another interpretation of the interrelationship between the general Part 3 rules and those relating to fraudulent adverts remains viable.

If advertisers *are* considered 'users', more content will be dealt with under the general Part 3 rules (although not for search services; see the discussion earlier in section 2b), considerably narrowing the application of the ad fraud-specific duties. The impact of this interpretation on

fraudulent ads would not be great, given that fraud is also listed in Schedule 7 as priority content for the purposes of the general U2U duties. This classification would, however, mean that social media services would have duties – which came into force on 17 March 2025 - to address illegal advertisements they disseminate, not only those linked to the offences outlined in s 40. So, this would potentially require action against ads for firearms, illicit drugs, and deepfake pornography. Additionally, this interpretation suggests an obligation to prevent pornographic advertisements from reaching minors; an area where recent research indicates significant deficiencies.

Nonetheless, Ofcom's current approach does not appear to follow this interpretation, but seemingly adopts one in line with the policy intent. Questions have been raised during the consultations on the general codes. In the stakeholder response document for the illegal harms consultation, Ofcom noted (p. 68):

> *A1.7.7 The Global Disinformation Index wanted clarity as to whether programmatic advertising services would be within scope of these measures. It also highlighted that if these services are not captured there is potential risk that it does not report data that could reveal the extent to which illegal content is being funded and amplified on services.*
>
> *Our response:*
> *A1.7.8 The Act sets out the services that are in scope of the online safety regime. If a person provides an online service, it may be in scope of these duties. It is up to the provider to assess the nature of their service and, if necessary, seek independent specialist advice to determine whether or not their service would be subject to the requirements of the Act.*

This somewhat ambiguous response could be understood against a more recent elaboration in relation to its Code on Content Harmful to Children (Vol 2). Following concerns expressed by some respondents to the consultation about 'surveillance advertising', Ofcom commented (p. 44):

> *5. 49: […] We emphasise that all service providers must consider their revenue model in their children's risk assessments, as it is a general risk factor in the Children's Risk Profiles. The Children's Register details how revenue models, such as advertising to generate income, can create financial incentives that may lead businesses to expose children to harmful content. For instance, if harmful content is engaging, service providers may have a financial incentive to recommend such content to children in order to generate more revenue from advertising. Therefore, if a service provider uses 'surveillance advertising', it must consider the impact this has on children encountering*

harmful content on their service. However, it is beyond our powers under the Act to ban this form of advertising.

This seems to suggest that where adverts are linked to illegal or harmful content, then the business model and delivery system should be considered relevant for risk assessment and safety duties. Where the adverts are criminal or harmful, this in and of itself would not trigger the rules unless they constitute fraudulent ads. Non-content concerns about the targeted delivery of adverts would likewise not trigger the regime.

The inclusion of adverts in the regime implies that the business systems for accepting ad copy and delivery/ publishing the ads (insofar as it is operated by the provider as part of the service) would be included in the safety obligations. This would seem to be true as regards the specific fraudulent ad rules and - potentially - the general Part 3 rules, though the impact on those systems for general Part 3 rules purposes might seem more indirect than for the fraudulent ad rules. Moreover, external delivery services and other actions in the online advertising environment would escape this net, but they would still need to comply with e.g., data protection rules. In principle, this could mean that some problems around ad fraud remain unaddressed, despite these layered obligations.

Although the government at the time aimed (see section 2, p. 4) to confine regulatory duties for paid-for ads to a narrow set of fraud-related offences (Chapter 5, OSA), the legislation's text leaves open broader interpretations, particularly concerning whether certain ads fall within the general Part 3 illegal content duties. Crucially, these general duties became enforceable in March 2025, whereas Chapter 5 obligations will not be enforceable until Ofcom's Code of Practice takes effect, likely in 2026 (see section 2c, p. 7). If a broader reading of 'user-generated content' and 'user' prevails, U2U services may already be subject to wider obligations regarding illegal ads. This interpretative ambiguity allows for a divergence in enforcement expectations from what the previous government intended, potentially accelerating regulatory exposure for in-scope services before Ofcom's ad-specific rules come into force.

3. The Online Advertising Programme (OAP)

The OAP was intended to look at the online advertising market across the board. It is still short on delivering results. The then government formed a taskforce; more concretely, the Online Fraud Charter was agreed. This is a set of voluntary commitments for companies to sign on to. Although a six-month review was scheduled, there is no sign of an output from that review (assuming the review even took place, given the timeline was before the 2024 General Election). One year after its implementation, Which? research indicates that the Charter has had limited success in reducing scams: despite the commitments made, fraud is 'still rife' on

online platforms in the UK. The OAP taskforce released a progress report, but its focus appears to mirror that of the OSA, largely addressing content-based issues, rather than the broader advertising delivery system (despite the concerns noted in prior consultations above). This in parallel raises the question: if the OAP's scope aligns with the OSA, why were ads excluded in the first place? The Labour Government has confirmed the OAP will continue for 12 months from December 2024 under updated terms of reference that reflect this focus.

4. Conclusions

While the OSA introduces important duties for platforms to prevent, minimise, and remove fraudulent ads, several unresolved issues might limit its effectiveness. Although the statutory instruments underpinning the categorisation rules are now in force, 'small but risky' platforms have been excluded, so we know that the fraudulent ads provisions will only apply to large Category 1 or Category 2a services, even though these have not yet been named. The inconsistency in how duties apply to user-generated content *vs* paid-for advertising further complicates enforcement, particularly with ambiguities around the definition of 'users' and how this affects advertisers.

The exclusion of paid-for ads under the general Part 3 duties for search services, while seemingly including them for U2U services, creates a potential loophole. Notably, a wider interpretation of these general duties for U2U platforms could relieve some enforcement pressure from the ASA, which currently lacks formal leverage under the OSA (it might enable the ASA to act as a trusted flagger or take on a more integrated enforcement role). This advantage, however, would not extend to search services, underscoring a further limitation of the current regime. This disparity suggests that platforms like social media may face stricter oversight than search engines, despite the similarity in the risks posed by illegal ads across both types of service. As noted, however, Ofcom proceeds on the basis that paid-for ads do not constitute user-generated content and are therefore exempt from the general Part 3 rules, which are now in effect and enforceable as of 17 March 2025. While this interpretation may be a practical approach to the issue, it also means that stronger user protections, which would have now taken effect had a broader reading been adopted, have not materialised.

Finally, the regime does not directly duplicate the ASA obligations but complements them by placing responsibility on platforms. However, it still leaves search services outside the scope of this potential benefit, and leaves room for external advertising delivery systems to escape oversight - a significant gap that OAP initially appeared poised to address but does not. Overall, while the OSA framework strengthens platform responsibilities, its relatively narrow scope, the uncertainty of its application across different services, and unaddressed external ad delivery processes leave some vulnerabilities.